Courant Mathematics and

Computing Laboratory

U. S. Department of Energy

High Resolution Schemes for Hyperbolic Conservation Laws

Ami Harten

Research and Development Report

Mathematics and Computing

March 1982

New York University

Courant Mathematics and Computing Laboratory

New York University

Mathematics and Computing   DOE/ER/03077-175

HIGH RESOLUTION SCHEMES FOR

HYPERBOLIC CONSERVATION LAWS

Ami Harten [*]

[*] School of Mathematical Sciences, Tel-Aviv University.

TABLE OF CONTENTS

Abstract.

This paper presents a class of new explicit second order accurate finite difference schemes for the computaiton of weak solutions of hyperbolic conservation laws. These highly nonlinear schemes are obtained by applying a nonoscillatory first order accurate scheme to an appropriately modified flux function. The so derived second order accurate schemes achieve high resolution while preserving the robustness of the original nonoscillatory first order accurate scheme.

Numerical experiments are presented to demonstrate the performance of these new schemes.

1. Introduction.

In this paper we consider numerical approximations to weak solutions of the initial value problem (IVP) for hyperbolic systems of conservation laws.

$$u_t + f(u)_x = 0, \quad u(x,0) = \phi(x), \quad -\infty < x < \infty. \tag{1.1}$$

Here $u(x,t)$ is a cloumn vector of m unknowns, and $f(u)$, the flux, is a vector valued function of m components. (1.1) is called hyperbolic if all eigenvalues $a_1(u),\ldots, a_m(u)$ of the Jacobian matrix $A(u)$

$$A(u) = f_u \tag{1.2a}$$

are real and the set of right eignevectors $R^1(u),\ldots, R^m(u)$ is complete. We assume that the eigenvalues $\{a^i(u)\}$ are arranged in a nondecreasing order

$$a^1(u) \le a^2(u) \le \ldots \le a^m(u). \tag{1.2b}$$

We consider systems of conservation laws (1.1) that possess an entropy function $U(u)$, defined as follows:

(i) U is a convex function of u, i.e., $U_{uu} > 0$,
(ii) U satisfies

$$U_u f_u = F_u \tag{1.3a}$$

where F is some other function called entropy flux. Admissible weak solutions of (1.1) satisfy, in the weak sense, the following inequality:

$$U(u)_t + F(u)_x \leq 0 \qquad\qquad (1.3b)$$

(see [11]). The inequality (1.3b) is called an entropy condition.

In the following we shall discuss numerical approximations to weak solutions of (1.1) which are obtained by (2k+ 1)-point explicit schemes in conservation form

$$v_j^{n+1} = v_j^n - \lambda \left(\bar{f}_{j+\frac{1}{2}}^n - \bar{f}_{j-\frac{1}{2}}^n \right)$$

where

$$\bar{f}_{j+\frac{1}{2}}^n = f(v_{j-k+1}^n, \ldots, v_{j+k}^n) \qquad\qquad (1.4b)$$

Here $v_j^n = v(j\Delta x, n\Delta t)$, an $\bar{f}$ is a numerical flux function. We require the numerical flux function to be consistent with the flux $f(u)$ in the following sense:

$$\bar{f}(u, \ldots, u) = f(u). \qquad\qquad (1.4c)$$

We say that the difference scheme (1.4) is consistent with the entropy condition (1.3b) if an inequality of the following kind is satisfied:

$$U_j^{n+1} \leq U_j^n - \lambda \left(\bar{F}_{j+\frac{1}{2}}^n - \bar{F}_{j-\frac{1}{2}}^n \right) \qquad\qquad (1.5a)$$

where $U_j^n = U(v_j^n)$, $\bar{F}_{j+\frac{1}{2}}^n = \bar{F}(v_{j-k+1}^n, \ldots, v_{j+k}^n)$; here $\bar{F}$ is a numerical entropy flux, consistent with the entropy flux $F(u)$, i.e.

$$\bar{F}(u,\ldots,u) = F(u). \tag{1.5b}$$

We turn now to discuss the question of convergence of the finite difference solution of (1.4) to weak solutions of the conservation laws (1.1). Since the finite-difference scheme is nonlinear and the computed solutions are certainly not smooth, therefore L_2-stability of a consistent finite-diference scheme does not imply convergence. One can establish convergence of finite - difference solutions of (1.4) to weak solutions of (1.1) when the following conditions are satisfied:

(i)   The total variation with respect to x of the finite-difference solutions is uniformly bounded with respect to t, Δt and Δx.

(ii) The finite-difference scheme (1.4) is consistent with the entropy ocndition (1.3b) for all entropy functions of (1.1).

(iii) The entropy condition (1.3b) implies uniqueness of the solution to the IVP (1.1).

Using compactness arguments one can deduce from condition (i) the existence of convergent subsequences. The conservation form (1.4) and condition (ii) imply that each limit solution is a weak solution which satisfies the entropy condition (1.3b). When the entropy condition implies uniqueness of the IVP (conditon (iii)) then all subsequences have the same limit solution, and con-sequently the finite-difference scheme is convergent.(see [2],[9], [10]).

It seems possible to satisfy conditions (i) and (ii) by adding a hefty amount of artificial viscosity to the finite difference scheme (1.4). The addi-tional viscosity terms damp possible oscillations in the computed solution,and

make the convergence process simulate the zero-dissipation-limit which is used to select the unique physically relevant weak solution. Unfortunately, viscosity represents an irretrievable loss of information and therefore the addition of artificial viscosity brings about some deterioration in resolution.

In this paper we describe a new method to design finite-difference schemes that satisfy conditions (i) and (ii), but are second-order accurate and have high resolution.

2. Monotonicity in the scalar case.

In this section we consider the IVP for a scalar conservation law.

$$u_t + f(u)_x \equiv u_t + a(u)u_x = 0, \quad a(u) = \frac{df}{du} \tag{2.1a}$$

$$u(x,0) = \phi(x), \qquad\qquad -\infty < x < \infty, \tag{2.1b}$$

where $\phi(x)$ is assumed to be of bounded total variation.   Every

weak solution of the scalar IVP (2.1) which satisfies the entropy

condition has the following <u>monotonicity property</u> as a function of t:

i)   No new local extrema in x may be created.

ii) The value of a local minimum is nondecreasing, the value of a local maximum

is nonincreasing.

It follows from this monotonicity property that the total variation in

x, $TV(u(t))$, of $u(x,t)$, is nonincreasing in t, i.e.,

$$TV(u(t_2)) \leqslant TV(u(t_1)), \quad \text{for all } t_2 \geqslant t_1. \tag{2.2}$$

We consider now explicit $(2k + 1)$ point finite difference schemes in con-

servation form (1.4) approximating (2.1)

$$v_j^{n+1} = H(v_{j-k}^n, v_{j-k+1}^n, \ldots, v_{j+k}^n) = v_j^n - \lambda[\bar{f}(v_{j-k+1}^n, \ldots, v_{j+k}^n) \tag{2.3a}$$

$$- \bar{f}(v_{j-k}^n, \ldots, v_{j+k-1}^n)].$$

and denote (2.3a) in an operator form as

$$v^{n+1} = L \cdot v^n \tag{2.3b}$$

We say that the finite difference scheme (2.3) is <u>total variation nonincreas-</u> <u>ing (TVNI)</u> if for all v of bounded total variation

$$TV(L \cdot v) \leqslant TV(v) \tag{2.4a}$$

where

$$TV(u) \equiv \sum_{j=-\infty}^{\infty} |\Delta_{j+1/2} u| \quad ; \tag{2.4b}$$

here, and throughout this paper, we use the standard notation

$$\Delta_{j+1/2} u = u_{j+1} - u_j, \tag{2.5}$$

We say that the finite-difference scheme (2.3) is <u>monotonicity preserving</u> if the finite difference operator L is monotonicity preserving, i.e., if v is a monotone mesh function so is $L \cdot v$.

We say that the finite difference scheme (2.3) is a <u>monotone scheme</u> if H in (2.3a) is a monotone nondecreasing function of each of its $2k + 1$ arguments.

The following theorem states the hierarchy of these properties.

<u>Theorem 2.1.</u> (i) A monotone scheme is TVNI. (ii) A TVNI scheme is monotonicity preserving.

Proof: (i) It was proven by B. Keyfitz in [8] that monotone schemes form an ℓ_1-contractive semigroup, i.e.

$$||L \cdot v - L \cdot z||_{\ell_1} \leq ||v - z||_{\ell_1} \tag{2.6a}$$

for all ℓ_1-summable v and z; Here $||u||_{\ell_1} = \sum_{j=-\infty}^{\infty} |u_j|$. (2.4) follows immediately from applying (2.6a) to v and $z = T \cdot v$, (i.e. $z_j \equiv v_{j+1}$ for all j).

(ii) Let (2.3) be a TVNI scheme and let v be a monotone mesh function of bounded total variation, and denote $w = L \cdot v$. Since L has a finite support of $2k + 1$ points it is sufficient to prove that w is monotone for all v of the form

$$v = \begin{cases} \text{constant} = v_L & j \leq J_- \\ \text{monotone} & J_- \leq j \leq J_+ \\ \text{constant} = v_R & j \geq J_+ \end{cases} \qquad , \quad J_+ \geq J_- \quad .$$

$$TV(v) = |v_R - v_L| \tag{2.6b}$$

We prove (ii) by negation. Suppose w is not monotone, then it has at least one local minimum and one local maximum. Denote by v_m and v_M the values of the first two successive local extrema, then

$$TV(w) \geq |v_R - v_L| + |v_m - v_M| > TV(v),$$

which contradicts the assumption that the scheme is TVNI. This completes the proof of Theorem 2.1

Monotone schemes approximate solutions of the viscous modified equation

$$u_t + f(u)_x = \Delta t[\beta(u,\lambda) \, u_x]_x \quad , \quad \lambda = \frac{\Delta t}{\Delta x} \tag{2.7a}$$

$$\beta(u,\lambda) = \frac{1}{2\lambda^2}\left[\sum_{\ell=-k} \ell^2 H_\ell(u,u,\ldots,u) - \lambda^2 a^2(u)\right] \tag{2.7b}$$

$$\beta(u,\lambda) \geq 0, \quad \beta(u,\lambda) \not\equiv 0 \tag{2.7c}$$

to second order accuracy; Since $\beta(u \; \lambda) \not\equiv 0$, monotone schemes are necessarily first order accurate; H_ℓ in (2.7c) denotes $\frac{\partial H}{\partial w_\ell}(w_{-k}, \, w_{-k+1}, \ldots, w_k)$. (see [8])

Since monotone schemes are TVNI, there exist convergent subsequences for all initial data of bounded total variation. Each limit is a weak solution of (2.1) that satisfies Oleinik's entropy condition (see [8]). Since Oleinik's entropy condition implies uniqueness of the IVP (2.1), we conclude that all subsequences converge to the same limit, and therefore the scheme is convergent (see [2]).

Let us consider now the scalar constant coefficient case $a(u) = $ constant in (2.1). A linear finite-difference approximation

$$v_j^{n+1} = \sum_{\ell=-k}^{k} c_\ell v_{j+\ell}^n \quad , \quad c_\ell = \text{const.} \tag{2.8a}$$

is monotonicity preserving if and only if

$$c_\ell \geq 0 \; , \quad\quad\quad -k \leq \ell \leq k.$$

(see [4]). Hence any _linear_ monotonicity preserving scheme, and therefore any TVNI linear scheme, is a monotone scheme, and consequently first order accurate.

We remark that the previous statement does not exclude the possibility of having _nonlinear_ monotonicity preserving and TVNI schemes that are second order accurate (and consequently are not monotone schemes). In fact the schemes presented in [7] and [6] are monotonicity preserving (at least in the constant coefficient case) and second order accurate.

It is the purpose of this paper to present new high-resolution second order accurate TVNI schemes. These new schemes are generated by converting known 3-point first order accurate TVNI schemes into new 5-point second order accurate TVNI schemes. Both the 3-point schemes and the new 5-point schemes can be re-written in the form

$$v^{n+1} = L \cdot v^n \qquad (2.9a)$$

$$(L \cdot v)_j = v_j + C_{+,j+1/2} \Delta_{j+1/2} v - C_{-,j-1/2} \Delta_{j-1/2} v \qquad (2.9b)$$

where $\Delta_{j+\frac{1}{2}}$ is defined in (2.5) and

$$C_{+,j+1/2} = C_+(v_{j-1}, v_j, v_{j+1}, v_{j+2}), \quad C_{-,j-1/2} = C_-(v_{j-2}, v_{j-1}, v_j, v_{j+1}) \quad (2.9c)$$

The following Lemma states conditions on the coefficients (2.9c) which are sufficient to ensure that the scheme (2.9) is TVNI.

Lemma 2.2

Let the coefficients C in (2.9c) satisfy the inequalities

$$C_{-,j+1/2} \geq 0 \;, \quad C_{+,j+1/2} \geq 0 \tag{2.10a}$$

$$C_{-,j+1/2} + C_{+,j+1/2} \leq 1 \;; \tag{2.10b}$$

then the scheme (2.9) is TVNI.

Proof: Denote $w = L.v$ and subtract (2.9b) at $j = i$ from (2.9b) at $j = i+1$ to obtain

$$\Delta_{i+1/2}w = C_{-,i-1/2}\Delta_{i-1/2}v + (1-C_{-,i+1/2}-C_{+,i+1/2})\Delta_{i+1/2}v + C_{+,i+3/2}\,\Delta_{i+3/2}v. \tag{2.11}$$

By (2.10) all the coefficients in (2.11) are non negative, therefore

$$|\Delta_{i+1/2}w| \leq (1 - C_{-,i+1/2} - C_{+,i+1/2})|\Delta_{i+1/2}v| + C_{-,i-1/2}|\Delta_{i-1/2}v| \tag{2.12}$$

$$+ C_{+,i+3/2}|\Delta_{i+3/2}v|$$

Summing (2.12) for $-\infty < i < \infty$ we obtain

$$TV(w) \equiv \sum_{i=-\infty}^{\infty} |\Delta_{i+1/2}w| \leq \sum_{i=-\infty}^{\infty} (1 - C_{-,i+1/2} - C_{+,i+1/2})|\Delta_{i+1/2}v|$$

$$+ \sum_{i=-\infty}^{\infty} C_{-,i-1/2}|\Delta_{i-1/2}v| + \sum_{i=-\infty}^{\infty} C_{+,i+3/2}|\Delta_{i+3/2}v| = \sum_{i=-\infty}^{\infty} |\Delta_{i+1/2}v| \equiv TV(v)$$

which shows that (2.4) is satisfied. The equality is obtained by changing the summation index in the last two sums in the RHS of the inequality.

In the next section we shall use Lemma 2.2 to design second order accurate TVNI schemes. We remark that any 3 point finite difference scheme in conservation form with a differentiable numerical flux can be rewritten as (2.9), in the following way: It follows from the mean value theorem that there exist C_+ and C_- such that

$$\lambda[\bar{f}(v_j, v_{j+1}) - \bar{f}(v_j, v_j)] = -C_+(v_j, v_{j+1}) \, \Delta_{j+1/2} v \qquad (2.13a)$$

$$\lambda[\bar{f}(v_{j-1}, v_j) - \bar{f}(v_j, v_j)] = -\, C_-(v_{j-1}, v_j) \, \Delta_{j-1/2} v \qquad (2.13b)$$

Expressing the numerical flux values in (2.3a) with $k = 1$ by (2.13) results in the form (2.9).

3. <u>Second order accurate TVNI schemes</u>.

Let us consider a general 3-point finite difference scheme in conservation form (1.4) with a numerical flux $\bar{f}$ of the form

$$\bar{f}(v_j, v_{j+1}) = \frac{1}{2} [f(v_j) + f(v_{j+1}) - \frac{1}{\lambda} Q(\lambda \bar{a}_{j+\frac{1}{2}}) \Delta_{j+\frac{1}{2}} v] \qquad (3.1a)$$

where

$$\bar{a}_{j+\frac{1}{2}} \begin{cases} [f(\bar{v}_{j+1}) - f(v_j)]/\Delta_{j+\frac{1}{2}} v & \text{when } \Delta_{j+\frac{1}{2}} v \neq 0 \\ a(v_j) & \text{when } \Delta_{j+\frac{1}{2}} v = 0 \end{cases} \qquad (3.1b)$$

Here $Q(x)$ is some function, which is often referred to as the coefficient of numerical viscosity.

<u>Lemma 3.1</u> Let $Q(x)$ in (3.1a) satisfy the inequalities

$$|x| \leq Q(x) \leq 1 \qquad \text{for } 0 \leq |x| \leq \mu \leq 1, \qquad (3.2)$$

then the finite-difference scheme (1.4) with (3.1) is TVNI under the CFL-like restriction.

$$\lambda \max_j |\bar{a}_{j+\frac{1}{2}}^n| \leq \mu \qquad (3.3)$$

<u>Proof</u>: Using the notation

$$\bar{v}_{j+\frac{1}{2}} = \lambda \bar{a}_{j+\frac{1}{2}} \qquad (3.4a)$$

where $\bar{a}_{j+\frac{1}{2}}$ is (3.1b), we rewrite (3.1a) as

$$\lambda \bar{f}_{j+\frac{1}{2}} = \lambda \bar{f}(v_j, v_{j+1}) = \lambda f(v_j) - \frac{1}{2}\left[-\bar{v}_{j+\frac{1}{2}} + Q(\bar{v}_{j+\frac{1}{2}})\right] \Delta_{j+\frac{1}{2}} v \qquad (3.4b)$$

and similarly

$$\lambda \bar{f}_{j-\frac{1}{2}} = \lambda \bar{f}(v_{j-1}, v_j) = \lambda f(v_j) - \frac{1}{2}\left[\bar{v}_{j-\frac{1}{2}} + Q(\bar{v}_{j-\frac{1}{2}})\right] \Delta_{j-\frac{1}{2}} v \quad . \qquad (3.4c)$$

Substituting (3.4) for numerical flux values in (1.4) we get the form (2.9)

$$v_j^{n+1} = v_j^n - \lambda(\bar{f}_{j+\frac{1}{2}} - \bar{f}_{j-\frac{1}{2}}) = v_j^n + \frac{1}{2}\left[Q(\bar{v}_{j+\frac{1}{2}}) - \bar{v}_{j+\frac{1}{2}}\right] \Delta_{j+\frac{1}{2}} v^n \qquad (3.5a)$$

$$- \frac{1}{2}\left[Q(\bar{v}_{j-\frac{1}{2}}) + \bar{v}_{j-\frac{1}{2}}\right] \Delta_{j-\frac{1}{2}} v^n \equiv v_j^n + C_{+,j+\frac{1}{2}} \Delta_{j+\frac{1}{2}} v^n - C_{-,j-\frac{1}{2}} \Delta_{j-\frac{1}{2}} v^n.$$

where

$$C_{\pm,j+\frac{1}{2}} = \frac{1}{2}\left[Q(\bar{v}_{j+\frac{1}{2}}) \pm \bar{v}_{j+\frac{1}{2}}\right]. \qquad (3.5b)$$

Since

$$C_{+,j+\frac{1}{2}} + C_{-,j+\frac{1}{2}} = Q(\bar{v}_{j+\frac{1}{2}}) \qquad (3.5c)$$

it follows from (3.2) and (3.5) that conditions (2.10) of Lemma 2.2 are satisfied under the CFL restriction (3.3) and therefore the finite-difference scheme (3.1) is TVNI.

The second order accurate Lax-Wendroff scheme has the numerical flux (3.1) with $Q(x) = x^2$, i.e.,

$$\overline{f}_{j+1/2}^{LW} \equiv \frac{1}{2} \ [f(v_j) + f(v_{j+1}) - \frac{1}{\lambda} \ (\overline{v}_{j+1/2})^2 \Delta_{j+1/2} v]. \tag{3.6a}$$

Clearly a numerical flux of a second order accurate scheme $\overline{f}_{j+1/2}$ has to satisfy

$$\overline{f}_{j+1/2} - \overline{f}_{j+1/2}^{LW} = O(\Delta^2), \tag{3.6b}$$

for all smooth solutions of (1.1); here Δ is the discretization parameter. When $Q(x)$ is constrained by (3.2), then the 3-point scheme (3.1) is only first order accurate, for

$$\overline{f}_{j+1/2} = \overline{f}_{j+1/2}^{LW} - \frac{1}{2\lambda} \ [-(\overline{v}_{j+1/2})^2 + Q(\overline{v}_{j+1/2})] \ \Delta_{j+1/2} \ v^n \tag{3.7}$$

and therefore

$$|f_{j+1/2}^{LW} - \overline{f}_{j+1/2}| \geq \frac{1}{2\lambda} \ [|\overline{v}_{j+1/2}| - (\overline{v}_{j+1/2})^2] \cdot |\Delta_{j+1/2} \ v^n| = O(\Delta).$$

We describe now how to convert a 3-point first order accurate TVNI scheme to a 5-point second order accurate TVNI scheme. Consider the application of a 3-point first order accurate TVNI scheme (3.1) to modified mesh values f_i^M of the original flux $f(u)$: Set

$$f_i^M = f(v_i) + \frac{1}{\lambda} \ g_i \ , \qquad\qquad g_i = g(v_{i-1}, v_i, v_{i+1}) \tag{3.8a}$$

where g is a function which will be specified below. The modified numerical flux $\overline{f}_{j+1/2}^M = \overline{f}^M(v_{j-1}, v_j, v_{j+1}, v_{j+2})$ is obtained by replacing $f(v_i)$, $f(v_{i+1})$ in (3.1a), (3.1b), by the modified values:

$$\bar{f}^M_{j+1/2} = \frac{1}{2} \{f^M_j + f^M_{j+1} - \frac{1}{\lambda} Q(\bar{\nu}^M_{j+1/2})\Delta_{j+1/2} v\} \tag{3.8b}$$

where

$$\bar{\nu}^M_{i+1/2} = \bar{\nu}_{i+1/2} + \gamma_{i+1/2} \ , \qquad \gamma_{i+1/2} = (g_{i+1} - g_i)/\Delta_{i+1/2} v. \tag{3.8c}$$

We can rewrite (3.8b) as

$$\bar{f}^M_{j+1/2} = \frac{1}{2} [f(v_j) + f(v_{j+1})] + \frac{1}{2\lambda} [g_j + g_{j+1} - Q(\bar{\nu}_{j+1/2} + \gamma_{j+1/2}) \Delta_{j+1/2} v]. \tag{3.8d}$$

Lemma 3.2 Suppose $Q(x)$ is Lipschitz continuous and g_j satisfies

$$g_j + g_{j+1} = [Q(\bar{\nu}_{j+1/2}) - (\bar{\nu}_{j+1/2})^2] \Delta_{j+1/2} v + O(\Delta^2) \tag{3.9a}$$

$$\gamma_{j+1/2} \cdot \Delta_{j+1/2} v \equiv g_{j+1} - g_j = O(\Delta^2) \tag{3.9b}$$

then the numerical flux (3.8) satisfies (3.6)

Proof: The modified numerical flux $\bar{f}^M_{j+1/2}$ (3.8d) differs from the original

flux $\bar{f}_{j+1/2}$ (3.1) in the following way :

$$\bar{f}^M_{j+1/2} = \bar{f}_{j+1/2} + \frac{1}{\lambda} \{g_j + g_{j+1} + [Q(\bar{\nu}_{j+1})$$
$$\tag{3.10a}$$
$$- Q(\bar{\nu}_{j+1/2} + \gamma_{j+1/2})] \Delta_{j+1/2} v\} .$$

Substituting (3.7) for $\bar{f}_{j+1/2}$ in (3.10a) we get that (3.6) holds if the
relation

$$(g_j + g_{j+1}) + [Q(\bar{\nu}_{j+1/2}) - Q(\bar{\nu}_{j+1/2} + \gamma_{j+1/2})] \Delta_{j+1/2} v =$$
$$\tag{3.10b}$$
$$[Q(\bar{\nu}_{j+1/2}) - (\bar{\nu}_{j+1/2})^2] \Delta_{j+1/2} v + O(\Delta^2)$$

is satisfied.

Since $Q(x)$ is Lipschitz continuous

$$|Q(\bar{v}_{j+1/2}) - Q(\bar{v}_{j+1/2} + \gamma_{j+1/2})| \le \text{const.}|\gamma_{j+1/2}|, \qquad (3.10c)$$

therefore if follows from (3.9b) that the second term on the LHS of (3.10b) is itself $O(\Delta^2)$; consequently (3.9a) implies (3.10b). This completes the proof of Lemma 3.2.

We construct $g_i = g(v_{i-1}, v_i, v_{i+1})$ that satisfies (3.9) in the following way:

$$g_i = s_{i+1/2}\max[0,\min(|\tilde{g}_{i+1/2}|, \tilde{g}_{i-1/2}\cdot s_{i+1/2})] \qquad (3.11a)$$

$$= \begin{cases} s_{i+1/2}\cdot\min(|\tilde{g}_{i+1/2}|, |\tilde{g}_{i-1/2}|) & \text{when } \tilde{g}_{i+1/2}\cdot\tilde{g}_{i-1/2} \ge 0 \\ 0 & \text{when } \tilde{g}_{i+1/2}\cdot\tilde{g}_{i-1/2} \le 0 \end{cases}$$

where

$$\tilde{g}_{i+1/2} = \frac{1}{2} [Q(\bar{v}_{i+1/2}) - (\bar{v}_{1+1/2})^2] \Delta_{i+1/2} v \qquad (3.11b)$$

$$s_{i+1/2} = \text{sign} (\tilde{g}_{i+1/2}). \qquad (3.11c)$$

Lemma 3.3: Let g_i be defined by (3.11),then relations (3.9a) and (3.9b) are satisfied, and

$$|\gamma_{j+1/2}| = |g_{i+1} - g_j|/|\Delta_{j+1/2}v| \le \frac{1}{2}| Q(\bar{v}_{j+1/2}) - (\bar{v}_{j+1/2})^2|. \qquad (3.12)$$

Proof: First let us assume that $\tilde{g}_{j+1/2} \cdot \tilde{g}_{j-1/2} \geq 0$, then using the definition (3.11a) and the relation $\min(a,b) = \frac{1}{2}[(a+b) - |a-b|]$ we get

$$g_j = \frac{1}{2}[\tilde{g}_{j-1/2} + \tilde{g}_{j+1/2} - s_{j+1/2}|\tilde{g}_{j+1/2} - \tilde{g}_{j-1/2}|]$$

(3.13a)

$$= \tilde{g}_{j\pm1/2} + \frac{1}{2}[\mp(\tilde{g}_{j+1/2} - \tilde{g}_{j-1/2}) - s_{j+1/2}|\tilde{g}_{j+1/2} - \tilde{g}_{j-1/2}|].$$

From (3.11b) we conclude that if v is smooth and $Q(x)$ at least Lipschitz continuous then

$$\tilde{g}_{j+1/2} - \tilde{g}_{j-1/2} = O(\Delta^2).$$

Thus (3.13a) and (3.13b) imply that

$$g_j = \tilde{g}_{j\pm1/2} + O(\Delta^2).$$

(3.13c)

It is easy to see that (3.13c) holds even if $\tilde{g}_{j-1/2} \cdot \tilde{g}_{j+1/2} \leq 0$, for then $\tilde{g}_j = 0$ but $\tilde{g}_{j\pm1/2} = O(\Delta^2)$ itself (since $\Delta_{j\pm1/2} = O(\Delta^2)$).

Relations (3.9a) - (3.9b) follow immediately by rewriting (3.13c) as

$$g_j = \tilde{g}_{j+1/2} + O(\Delta^2) , \qquad g_{j+1} = \tilde{g}_{j+1/2} + O(\Delta^2).$$

We turn now to prove (3.12). We observe from the definition (3.11a) that g_j and g_{j+1} cannot be of different sign, hence

$$|g_{j+1} - g_j| \leq \max(|g_j|, |g_{j+1}|)$$

$$\leq \max[\min(|\tilde{g}_{j-1/2}|, |\tilde{g}_{j+1/2}|), \min(|\tilde{g}_{j+1/2}|, |\tilde{g}_{j+3/2}|)]$$

$$\leq |\tilde{g}_{j+1/2}|.$$

Thus it follows from (3.11b) that

$$|\gamma_{j+1/2}| = |g_{j+1} - g_j| / |\Delta_{j+1/2} v| \leq |\tilde{g}_{j+1/2}| / |\Delta_{j+1/2} v|$$

$$= |Q(\bar{v}_{j+1/2}) - (\bar{v}_{j+1/2})^2| \quad ;$$

this completes the proof of (3.12).

We show now that the 5-point second order accurate scheme (3.8) with (3.11) is TVNI under the same CFL restriction of the original 3-point first order accurate TVNI scheme (3.1).

Lemma 3.4: Suppose $Q(x)$ satisfies (3.2) and g_j is defined by (3.11), then the finite-difference scheme (1.4) with the numerical flux (3.8) is TVNI under the CFL restriction (3.3).

Proof: Since (3.8) is (3.1) applied to a modified flux f_i^M (3.8a), it can be rewritten as (3.5) with $\bar{v}_{j+1/2}$ replaced by $\bar{v}_{j+1/2}^M$ (3.8b). We conclude from Lemma 3.1 that the scheme (3.8) is TVNI under the modified CFL restriction

$$\max_j |\bar{v}^M_{j+1/2}| \le \mu \qquad (3.14)$$

To complete the proof of Lemma 3.4 we show that (3.14) is implied by the original CFL condition (3.3). Using (3.12) and (3.2) we get

$$|\bar{v}^M_{j+1/2}| = |\bar{v}_{j+1/2} + \gamma_{j+1/2}| \le |\bar{v}_{j+1/2}| + |\gamma_{j+1/2}| \le |\bar{v}_{j+1/2}| +$$

$$+ \frac{1}{2} |Q(\bar{v}_{j+1/2}) - (\bar{v}_{j+1/2})^2| \le |\bar{v}_{j+1/2}| + \frac{1}{2} [1 - (\bar{v}_{j+1/2})^2] =$$

$$= 1 - \frac{1}{2} (|\bar{v}_{j+1/2}| - 1)^2 \le 1$$

whenever $|\bar{v}_{j+1/2}| \le Q(\bar{v}_{j+1/2}) \le 1$; this shows that (3.14) holds.

Remarks: (1) If $\Delta_{j+1/2}v = 0$ then it follows immediately from the definition (3.11) that $g_j = g_{j+1} = 0$. This shows that the modified numerical flux (3.8d) is consistent with the physical flux $f(u)$ in the sense of (1.4c). The scheme (3.8) + (3.11) is TVNI and therefore it has convergent subsequences for all initial data of bounded total variation; the limits of these subsequences are weak solutions of the scalar conservation law (2.1). To complete the convergence proof one has to show that all these limits are the same. In the constant coefficient case the solution to the IVP (2.1) is unique and therefore the scheme is convergent (note that the scheme is nonlinear even in the constant coefficient case!) In the nonlinear case, convergence will follow if one shows that the scheme is consistent with Oleinik's entropy condition in the sense of (1.5). We shall discuss consistency with the entropy condition in Section 5.

(2) Condition (3.6) is only a necessary condition for second order accuracy.
It becomes a sufficient condition if the coefficient in the $O(\Delta^2)$ term in (3.6)
is differentiable, except possibly at a finite number of points $N(t, \Delta t)$, such
that $N\Delta t \to 0$ as $\Delta t \to 0$ for all t. It is clear from (3.13) and (3.9) that
the troublesome points where the scheme (3.8) + (3.11) may degenerate locally to
$O(\Delta^2)$ truncation error are those where $s_{j+1/2}$ (3.11c) is discontinuous, i.e.
where $Q(\nu) - \nu^2 = 0$ or $u_x = 0$. The fact that the scheme is TVNI controls
the possible increase of the number of local extremum points in the computed
solution. The schemes that we consider in section 5 all have the monotonicity
property (see section 2); i.e. the number of local extremum points in the computed
solution is nonincreasing in time, and thus bounded by that of the initial data.

(3) The modified equation of the scheme (3.1), i.e., the equation which it approxi-
mates to second order accuracy, is (2.7) with

$$\beta(u,\lambda) = \left[Q(\nu) - \overline{\nu^2}\right]\frac{1}{2\lambda^2}, \qquad \nu = \lambda a(u) \tag{3.15a}$$

We rewrite the modified equation as

$$u_t + \{f - \frac{\Delta t}{2\lambda^2}[Q(\nu) - \nu^2] u_x\}_x = 0, \tag{3.15b}$$

and observe that $\overset{\smile}{g}_{i+1/2}$ (3.11b), and consequently g_i (3.11a), is an
approximation to the term

$$g \approx \frac{\Delta t}{2\lambda}[Q(\nu) - \nu^2] u_x = \frac{1}{2} [Q(\nu) - \nu^2] (\Delta x\, u_x). \tag{3.15c}$$

Our method to convert a first order accurate TVNI scheme into a second order accurate TVNI scheme is based on the following heuristic argument: The first order scheme (3.1) approximates

$$u_t + [f - \frac{1}{\lambda} g]_x = 0 \tag{3.16a}$$

to second order accuracy. Therefore, applying the same scheme to

$$u_t + [f + \frac{1}{\lambda} g]_x = 0 \tag{3.16b}$$

results in a second order accurate approximation to $u_t + f_x = 0$. To be able to apply the scheme to (3.16b) we have to define $g(u)$ with a bounded derivative dg/du, therefore $\tilde{g}_{i+1/2}$ (3.11b) is replaced by g_i (3.11a) (see [5]).

4.   Systems of Conservation Laws.

In this section we describe how to extend our new scalar scheme of section 3 to systems of conservation laws. Our extension technique is a somewhat generalized version of the procedure suggested by P. Roe in [14]. The basic idea is to extend the scalar scheme to the system case by applyiny it"scalarly" to each of the (appropriately linearized) characteristic variables.

$$\text{Let } S(u) = (R^1(u), R^2(u),\ldots,R^m(u)) \tag{4.1a}$$

be a matrix, the columns of which are the right eigenvectors of the Jacobian matrix $A(u)$ (1.2a). Then

$$S^{-1}AS = \Lambda \quad , \qquad \Lambda_{ij} = a^i(u)\delta_{ij} \quad . \tag{4.1b}$$

The rows $L^1(u)$, $L^2(u),\ldots,L^m(u)$ of $S^{-1}(u)$ constitute a complete system of left eigenvectors of $A(u)$ which is bi-orthonormal to the system of right eigenvectors, i.e.

$$L^i R^j = \delta_{ij}. \tag{4.1c}$$

In the constant coefficient case $A(u) \equiv A = \text{const.}$

$$u_t + Au_x = 0, \quad u(x,0) = \phi(x), \qquad -\infty < x < \infty. \tag{4.2}$$

one defines characteristic variables $w = (w^k)$ by

$$w^k = L^k u, \qquad w = S^{-1}u. \tag{4.3}$$

It follows from (4.1) that (4.2) decouples into m scalar characteristic equations, $1 \leq k \leq m$

$$w_t^k + a^k w_x^k = 0, \qquad w^k(x,0) = L^k \phi(x), \qquad -\infty < x < \infty. \qquad (4.4)$$

This offers a natural way of extending a scalar scheme to a constant coefficient system of equations (4.2) by applying it "scalarly" to each of the m scalar characteristic equations (4.4).

The characteristic variables w^k in (4.3) can also be viewed as the components of u in the coordinate system $\{R^k\}$, i.e.

$$u = \sum_{k=1}^{m} w^k R^k \qquad (4.5)$$

We use this interpretation of characteristic variables to extend the scalar scheme to general nonlinear systems of conservation laws.

Let $v_{j+1/2} = V(v_j, v_{j+1})$ be an average of v_j and v_{j+1}, i.e. a smooth function $V(u,v)$ such that

$$V(u,v) = V(v,u) \qquad (4.6a)$$

$$V(u,u) = u \qquad (4.6b)$$

and let $\alpha_{j+1/2}^k$ denote the component of $\Delta_{j+1/2} v = v_{j+1} - v_j$ in the coordinate system $\{R^k(v_{j+1/2})\}$

$$\Delta_{j+1/2} v = \sum_{k=1}^m \alpha_{j+1/2}^k R_{j+1/2}^k \tag{4.7a}$$

$$\alpha_{j+1/2}^k = L_{j+1/2}^k \Delta_{j+1/2} v ; \tag{4.7b}$$

here we use the notation convention $b_{j+1/2} = b(v_{j+1/2}) = b(V(v_j, v_{j+1}))$.

We now extend the scalar scheme $(3.8) + (3.11)$ to general systems of conservation laws as follows:

$$v_j^{n+1} = v_j^n - \lambda(\bar{f}_{j+1/2}^n - \bar{f}_{j-1/2}^n), \tag{4.8a}$$

$$\bar{f}_{j+1/2} = \frac{1}{2} [f(v_j) + f(v_{j+1})] +$$
$$\tag{4.8b}$$
$$+ \frac{1}{2\lambda} \sum_{k=1}^m R_{j+1/2}^k [g_j^k + g_{j+1}^k - Q^k(\nu_{j+1/2}^k + \gamma_{j+1/2}^k) \alpha_{j+1/2}^k]$$

where $\nu_{j+1/2}^k = \lambda a^k(v_{j+1/2})$ and

$$g_i^k = s_{i+1/2}^k \max[0, \min(|\tilde{g}_{i+1/2}^k|, \tilde{g}_{i-1/2}^k \cdot s_{i+1/2}^k)], \quad s_{i+1/2}^k = \text{sign}(\tilde{g}_{i+1/2}^k) \tag{4.8c}$$

$$\tilde{g}_{i+1/2}^k = \frac{1}{2} [Q^k(\nu_{i+1/2}^k) - (\nu_{i+1/2}^k)^2] \alpha_{i+1/2}^k \tag{4.8d}$$

$$\gamma_{i+1/2}^k = \begin{cases} (g_{i+1}^k - g_i^k)/\alpha_{i+1/2}^k & \text{when} \quad \alpha_{i+1/2}^k \neq 0 \\ 0 & \text{when} \quad \alpha_{i+1/2}^k = 0 \end{cases} \tag{4.8e}$$

The second order accurate one-step Lax-Wendroff scheme can be represented
as

$$\bar{f}_{j+1/2}^{LW} = \frac{1}{2} \left[f(v_j) + f(v_{j+1}) \right] - \frac{1}{2\lambda} \sum_{k=1}^{m} (v_{j+1/2}^k)^2 \alpha_{j+1/2}^k R_{j+1/2}^k \qquad (4.9)$$

$$\equiv \frac{1}{2} \left[f(v_j) + f(v_{j+1}) \right] - \frac{\lambda}{2} \left[A(v_{j+1/2}) \right]^2 \Delta_{j+1/2} v$$

Lemma 4.1. Suppose $\{Q^k(x)\}$ are Lipschitz continuous, then (4.8) satisfies

$$\bar{f}_{j+1/2} = \bar{f}_{j+1/2}^{LW} + O(\Delta^2) \qquad (4.10)$$

Proof: Rewrite (4.8b) as

$$\bar{f}_{j+1/2} = f_{j+1/2}^{-LW} + \sum_{k=1}^{m} R_{j+1/2}^k \tau_{j+1/2}^k \qquad (4.11a)$$

where

$$\tau_{j+1/2}^k = \frac{1}{2\lambda} \{ \varepsilon_j^k + \varepsilon_{j+1}^k - [Q^k(v_{j+1/2}^k) - (v_{j+1/2}^k)^2] \alpha_{j+1/2}^k \}$$

$$(4.11b)$$

$$- \frac{1}{2\lambda} [Q^k(v_{j+1/2}^k + \gamma_{j+1/2}^k) - Q^k(v_{j+1/2}^k)] \alpha_{j+1/2}^k ,$$

and then use (3.10c) and conclude (3.9) from Lemma 3.3.

We define the total variation $TV(v)$ of the vector mesh function v to
be

$$TV(v) = \sum_{j=-\infty}^{\infty} \sum_{k=1}^{m} |\alpha_{j+1/2}^k| \qquad (4.12)$$

where $\alpha_{j+1/2}^k$ is defined by (4.7), and show:

Lemma 4.2 Suppose $Q^k(x)$ satisfies (3.2) for all k, and that $A(u) \equiv A = $ constant
then the scheme (4.8) is TVNI under the CFL restriction

$$\lambda\max|A^k| \leqslant \mu = \min \mu^k \leqslant 1 \qquad\qquad (4.13)$$

where μ^k are the restrictions in (3.2).

Proof: Because of the assumption $A(u) \equiv$ const, $\{R^k\}$, $\{L^k\}$ and $\{a^k\}$ are all constant. Multiplying (4.8b) from the left by L^k, we obtain (3.8d) for the characteristic variable w^k (4.3); g_j^k and $\gamma_{j+1/2}^k$ in (4.8) becomes identical with (3.11). Thus by Lemma 3.4 we conclude that under condition (4.3) the total variation of each of the characteristic variables is nonincreasing, and therefore the total variation (4.12) is nonincreasing as well.

Corollary 4.3. The scheme (4.8) in the constant coefficient case is convergent under the restriction (4.13) for all initial data of bounded total variation, and is second under accurate.

We remark that this corollary is not trivial since the scheme is highly nonlinear even in the constant coefficient case.

Our technique to extend scalar schemes to the system case does not require any particualr form of averaging $V(u,v)$ (4.6). Roe in [15] is using a specific form of averaging that on top of being mathematically pleasing, also enables the computational advantage of perfectly resolving stationary discontinuities.

In [7] we show that if the system of conservation laws (1.1) possesses an entropy function (1.3), then it is symmetrizable, and there exists a mean value Jacobian $A(u,v)$ such that

(i) $f(v) - f(u) = A(u,v) (v-u)$ (4.14a)

(ii) $A(u,u) = A(u)$ (4.14b)

(iii) $A(u,v)$ has real eigenvalues $\{a^k(u,v)\}_{k=1}^m$ and a complete set of right eigenvectors $\{R^k(u,v)\}_{k=1}^m$.

In the context of the scheme (4.8) Roe's extension technique is expressed by taking $A_{j+1/2}^k$ and $R_{j+1/2}^k$ in (4.7) - (4.8) to be the eignevalues and the right eigenvectors of the mean value Jacobian $A(v_j, v_{j+1})$ (4.14a), respectively. Thus if $\alpha_{j+1/2}^k$ are defined by (4.7a)

$$v_{j+1} - v_j = \sum_{k=1}^m \alpha_{j+1/2}^k R_{j+1/2}^k \qquad (4.15a)$$

then it follows from (4.14a) that

$$f(v_{j+1}) - f(v_j) = \sum_{k=1}^m \alpha_{j+1/2}^k a_{j+1/2}^k R_{j+1/2}^k. \qquad (4.15b)$$

The relation (4.15) makes the scheme (4.8) a more faithful extension of (3.8) in the sense that (4.8) for $m=1$ is identical with the scalar scheme (3.8).

We observe that if $f(u) = f(v)$ in (4.14a) then $v - u$ is a right eigenvector or $A(u,v)$, corresponding to a zero eignvalue. Hence in (4.15) $\alpha^k(u,v) = 0$ for $k \neq k_o$ and $a^{k_o}(u,v) = 0$, for some k_o. It is easy to see that if $Q^{k_o}(0) = 0$ in (4.8) then the stationary discontinuity

$$u(x,t) \equiv \phi(x) = \begin{cases} u & x < 0 \\ v & x > 0 \end{cases} , \quad f(u) = f(v)$$

is also a stationary solution of (4.8). (see [15]).

In the case of the Euler equations of gasdynamics, where the flux $f(u)$ is an homogeneous function of u of degree 1, it is possible to express $A(u,v)$ in (4.14) as

$$A(u,v) = A(V(u,v)). \tag{4.17}$$

This relatively simple function $V(u,v)$ (see [15]) will be described in Section 7.

Remarks: (1) Note that we use $Q^k(x)$, thus allowing different functions (3.2) for different characteristic fields. As observed by P. Roe [14] the extension technique of this section permits even the use of completely different scalar schemes for different characteristic fields.

(2) In most applications it may be advantagous to replace the term $\frac{1}{2} [f(v_j) + f(v_{j+1})]$ in (4.8b) by $f(v_{j+1/2}) = f(V(v_j, v_{j+1}))$. For sure this simplifies the programming and reduces the CPU time, without altering the main properties of the scheme. A possible disadvantage of such a change is that the scheme may lose the property of perfect resolution of stationary discontinuities.

(3) The particular definition (4.12) of total variation is motivated by the definition of Glimm's functional in [3]. When applied to a piecewise-smooth solution $u(x,t)$ of (1.1)

$$\lim_{\Delta x \to 0} TV(u) = \int_{-\infty}^{\infty} \sum_{k=1}^{m} |L^k(u)u_x| dx + \sum_{j} \sum_{k=1}^{m} |\alpha^k(x_j)| \tag{4.18}$$

where x_j are points of discontinuity, and $\alpha^k(x_j)$ denotes the value of $\alpha^k_{i+1/2}$ in (4.7) evaluated with respect to $v_i = u((x_j)_-,t)$, $v_{i+1} = u((x_j)_+,t)$.

There is no reason to expect that the functional (4.18), and consequently (4.12), is generally nonincreasing with t. Based on Glimm's results [3] we do however believe that this functional (under certain conditions) is bounded in t. At this time we do not have estimates of the possible increase in total variation in solutions of the scheme (4.8), and therefore cannot prove convergence in the nonlinear system case.

5. On the choice of $Q(x)$.

In section 3 we have presented the basic idea of our new scheme in the scalar case, which can be described algorithmically by: "Take a 3-point TVNI scheme (3.1), which is first order accurate, and apply it to the modified flux value (3.8) to obtain a 5-point second order accurate TVNI scheme (3.11) + (3.8)

We begin section 5 by considering different choices of $Q(x)$ in the 3-point scheme (3.1). In the end of this section we argue that most of the properties of the 3-point point scheme (3.1) go over to the modified 5-point scheme (3.8) +(3.11).

In section 6 we discuss the system case.

A natural choice of $Q(x)$ in (3.2) is $Q(x) = |x|$, which corresponds to the least dissipative TVNI scheme of the form (3.1). The scheme (3.5) with $Q(x) = |x|$ can be rewritten as

$$v_j^{n+1} = v_j^n - (\bar{v}_{j+1/2})^- \Delta_{j+1/2}v^n - (\bar{v}_{j-1/2})^+ \Delta_{j-1/2}v^n \qquad (5.1a)$$

where

$$v^- = \min(v,0) = \frac{1}{2}(v - |v|), \quad v^+ = \max(v,0) = \frac{1}{2}(v + |v|). \qquad (5.1b)$$

This scheme (5.1) is a generalization of the well-known upstream differencing scheme of Courant, Isaacson and Rees, and it is well investigated in the literature (see [5], [10] and the references cited there). We now give a brief review of its relevant properties.

Lemma 5.1. Solutions of the upstream differencing scheme (5.1), under the CFL restriction

$$\lambda \max_{u \in \Phi} |a(u)| \le 1, \qquad \Phi = \text{range of initial data}, \qquad (5.2)$$

have the monotonicity property stated at the beginning of Section 2.

Sketch of Proof: Let $W(x)$ be the piecewise-linear interpolant of v^n, i.e.

$$W(x_j) = v_j^n , \quad x_j = j\Delta x.$$

Then (5.1), under the CFL restriction (5.2), implies that there exists a monotone nondecreasing sequence $\{\hat{x}_j\}$ such that

$$v_j^{n+1} = W(\hat{x}_j), \qquad |\hat{x}_j - x_j| \le \Delta x$$

This shows that v^{n+1} has no more local extremum points then v^n, and that the value of a local minimum (maximum) is nondecreasing (nonincreasing), which is the assertion of Lemma 5.1.

We observe that the stationary jump discontinuity (4.16), admissible or inadmissible, is also a sationary solution of the upstream differecing scheme (5.1). On one hand this is a desirable property as it implies good resolution of stationary shocks; on the other hand it indicates that the scheme many select nonphysical weak solutions that do not satisfy the entropy condition (1.3b). This property is related to the fact that the viscosity term $\beta(u,\lambda)$ (3.15a) vanishes for $v = 0$, and it is common to all schemes (3.1) with $Q(0) = 0$.

To prevent entropy condition violation of this sort and make $Q(x)$ smoother at the same time, we modify $Q(x) = |x|$ near $x = 0$ to be

$$Q(x) = \begin{cases} \dfrac{x^2}{4\epsilon} + \epsilon & \text{for} \quad |x| \leqslant 2\epsilon \\ |x| & \text{for} \quad |x| \geqslant 2\epsilon \end{cases} \tag{5.3}$$

with say $\epsilon = 0.05$. (see Fig. 1)

This change has increased the amount of numerical viscosity for $|x| < 2\epsilon$ so that now $\beta(u,\lambda) > 0$ for $|v| < 1$. $\beta(u,\lambda)$ vanishes only for $v = 1$; which can be handled by taking $\mu < 1$, say $\mu = 0.95$, in the CFL restriction (3.3). Consequently the scheme (3.1) with $Q(x)$ (5.3), cannot have perfect resolution of the stationary shock (4.16).

Another possible choice of $Q(x)$ which corresponds to a scheme that has already been investigated is $Q(x) = x^2 + \frac{1}{4}$ (see [1] and [6]). This scheme is the Lax-Wendroff scheme modified by the addition of the linear viscosity term $\frac{1}{8}(v_{j+1} - 2v_j + v_{j-1})$. Since now $Q(x) - |x| = (|x| - \frac{1}{2})^2 \geqslant 0$, it is the requirement $Q(x) \leqslant 1$ in (3.2) which restricts the CFL condition (3.3) to $\mu = \sqrt{3}/2 \approx 0.866 < 1$ (see Fig. 1).

We turn now to discuss the second order accurate TVNI scheme (3.8) + (3.11). We note that the truncation error of this scheme is $O(\Delta^3)$, except possibley where $u_x = 0$, independent of the particular form of $Q(x)$. Thus the modified scheme

removes all $O(\Delta^2)$ errors due to numerical viscosity, except at the points where u_x goes through zero; there some viscosity is needed to ensure monotonicity. Consequently we expect the new scheme to have high resolution of shocks, stationary or moving, almost independently of the particular choice of $Q(x)$.

Based on these considerations it seems to us that (5.3) is the best choice of $Q(x)$ for (3.11)

Next we deal with the question whether properties of the scheme (3.1) go over to the modified scheme (3.8) + (3.11). We expect properties of the original scheme that can be expressed as a relation between $F(v^n)$ and $F(v^{n+1})$, where F is some functional that depends on mesh values only, to go over to the modified second order accurate scheme. We have seen that this is true for the TVNI property; we can also prove that for $Q(x) = |x|$ the modified scheme has the monotonicity property, and that in the stationary discontinuity case (4.16) it behaves the same way as the original scheme.

To support our conjecture in general, we note that the modified scheme is the original scheme applied to a modified flux $f + \frac{1}{\lambda} g$ (3.16b). Thus if g were strictly a function of u, our statement becomes trivially true. However g in (3.11), which is consistent with zero flux, may be cosidered $g(v^n)$ only where v^n is monotone. The particular form used in (3.11) is designed to match the different definitions of $g(v^n)$ in neighbouring monotone sections by setting $g = 0$ at local extremum points (see [5]). That this matching is smooth, is evident by the fact that $\gamma = \Delta g / \Delta u$ is well-defined everywhere, and is bounded in absolute value by (3.12).

Hopefully the piecewise definition of $g(v^n)$ does not spoil validity of such properties for the modified scheme.

Based on these arguments we form the conjecture that if the original scheme (3.1) is consistent with the entropy inequality (1.3b), so is the modified second order accurate scheme (3.8) + (3.11).

6. On resolution and entropy.

The extension technique of section 4 consists of applying scalar schemes, not necessarily the same, to characteristic fields. Therefore we have the possibility to custom fit the scheme to the computational needs of each characteristic field.

We consider here systems of conservation laws where the characteristic fields are either genuinely nonlinear $(a_u^k R^k \neq 0)$ or linearly degenerate $(a_u^k R^k \equiv 0,$ see [11]). The waves of a genuinely nonlinear field are either shocks or rarefaction waves, depending whether the characteristics are convergent or divergent. The waves of a linearly degenerate field are exclusively contact discontinuities.

First let us consider the latter case of a linearly degenerate characteristic field. It is well known that the Riemann invariants of this field, one of which is the characteristic speed, are continuous across a contact discontinuity. Therefore the propagation of a contact discontinuity is computationally equivalent to that of a scalar discontinuity moving with a constant characteristic speed. As remarked in section 3, the solution of the first order accurate scheme (3.1) is the same, up to second order terms, as the solution of the modified differential equation (3.15b) (or (3.16a)). The modification of the flux $f(u) = $ constant.u by the addition of the term $(-\frac{1}{\lambda} g)$ has the effect of making the characteristic field slightly divergent in a region of a discontinuity. Consequently the computed discontinuity is being spread at the rate of $\sqrt{n}$, where n is the number of time-steps taken; similarly, a p-th order accurate standard scheme spreads a contact discontinuity at the rate of $n^{1/(p+1)}$ (see [5]).

Our new second order accurate scheme is obtained by applying the original one
to the flux $f + \frac{1}{\lambda} g$ (3.16b). This has the effect of reducing the divergence of
the characteristics induced by the original scheme in a region of discontinuity,
and we expect the computed discountinuity to spread like $n^{1/3}$, rather than $n^{1/2}$.

To ensure that the computed contact discontinuity does not spread in time at
all, we want to make the effective characteristic field slightly convergent. To
accomplish that we increase the size of g in (3.16b) e.g. by multiplying the RHS
of (4.8c) by $(1 + \theta_i^k)$, i.e.

$$g_i^k = (1 + \theta_i^k) s_{i+1/2}^k \max[0, \min (|\tilde{g}_{i+1/2}^k|, \tilde{g}_{i-1/2}^k \cdot s_{i+1/2}^k)] \qquad (6.1a)$$

where

$$\theta_i^k = |\alpha_{i+1/2}^k - \alpha_{i-1/2}^k| / (|\alpha_{i+1/2}^k| + |\alpha_{i+1-2}^k|) ; \qquad (6.1b)$$

$\alpha_{i+1/2}^k$ is defined by (4.7).

In [6] we show that

$$0 \le \theta_i^k \le 1 \qquad (6.2a)$$

$$\theta_i^k = 0(\Delta x) \quad \text{in regions of smoothness.} \qquad (6.2b)$$

and that $\theta_i^k = 0(1)$ in regions of discontinuities.

It is easy to see that this change in (4.8c) modifies the numerical flux
(4.8b) by $0(\Delta^2)$ and thus does not spoil the second order accuracy of the scheme.
From (6.2a) and (3.12) it follows that now

$$|\gamma_{j+1/2}^k| = |g_{j+1}^k - g_j^k|/|\alpha_{j+1/2}^k| \leq |Q(\nu_{j+1/2}^k) - (\nu_{j+1/2}^k)^2| \qquad (6.3a)$$

Hence for $Q(x)$ (3.2) we get in Lemma 3.4

$$|\nu_{j+1/2}^k| + |\gamma_{j+1/2}^k| \leq |\nu_{j+1/2}^k| + Q(\nu_{j+1/2}^k) - (\nu_{j+1/2}^k)^2 \qquad (6.3b)$$

For $Q(x) = |x|$ we have that the RHS of (6.3b) is less or equal 1 for $|\nu_{j+1/2}^k| \leq 1$, and consequently the modification (6.1) does not alter the CFL condition for this field. The same is true for $Q(x)$ in (5.3) with $\varepsilon < 1/3$. (However for $Q(x) = x^2 + 1/4$ (6.3b) implies a more stringent restriction, namely $\max_j |\nu_{j+1/2}^k| \leq 3/4$).

We remark that there are no entropy considerations associated with a linearly degenerate characteristic field.

We turn now to the case of a genuinely nonlinear characteristic field, where the computational aspects to be considered are resolution of shocks and enforcement of the entropy condition.

Unlike contact discontinuities, shocks are formed and maintained by local convergence of the characteristic curves. The reduction in numerical viscosity due to the addition of (4.8c) is usually sufficient to ensure good resolution of shocks. When higher resolution is desired, one may employ a mechanism of the form (6.1) to enhance the local convergence of the characteristics in the shock region and thus improve its resolution (see [5]).

The computational aspect of the entropy condition may be demonstrated by considering a Riemann problem for (1.1) involving two nearby states u_L and u_R. Lax in [11] proves (assuming all fields to be genuinely nonlinear) that there exists a unique "physical" solution to this problem, which consists in general of both shocks and rarefaction waves. One can show, using exactly the same method of proof, that there also exists a weak solution consisting only of shocks[1], which may differ from the physical one. Whenever the first contains rarefaction waves, the latter has nonphysical "expansion shocks". Therefore to ensure that the scheme selects a physically relevant weak solution, consistency with the entropy condition, on top of consistency with (1.1), is required.

In section 5 we have argued that the scalar scheme (3.8) + (3.11) is consistent with Oleinik's entropy condition. The question arises now whether the system version (4.8) of this scheme will always select the physical weak solution. T.P.Liu [12] shows that if a weak solution of (1.1) contains only "admissible" discontinuities that satisfy a "scalar" Oleinik condition with respect to the Rankine-Hungoniot curve, then this solution also satisfies the entropy inequality (1.3b). Using this theory the above question can be reformulated as: Will the scheme (4.8) reject discontinuities that do not satisfy Oleinik's condition along the Rankine-Hugoriot curve? It seems possible to answer this question affirmatively for sufficiently weak shocks, and feasibly for a larger class of problems.

----

1 We remark that there is also a one-parameter family of states connecting u_L and u_R through rarefaction curves only (see [13]). However a "negative rare-faction wave" is not computationally realizable as it is a multi-valued function of x.

7.   Application to Euler equations of gasdynamics

In this section we describe the application of our new scheme (4.8) to the Euler equation of gasdynamics:

$$w_t + f(w)_x = 0 \tag{7.1a}$$

$$w = \begin{pmatrix} \rho \\ m \\ E \end{pmatrix} \quad , \quad f(w) = uw + \begin{pmatrix} 0 \\ p \\ pu \end{pmatrix} \tag{7.1b}$$

$$p = (\gamma - 1)(E - 1/2 \rho u^2). \tag{7.1c}$$

Here ρ, u, p and E are the density, velocity, pressure and total energy, respectively; $m = \rho u$ is the momentum and we take $\gamma = 1.4$.

The eigenvalues of the Jacobian matrix $A(w) = f_w$ are

$$a_1(w) = u - c \quad , \quad a_2(w) = u \quad , \quad a_3(w) = u + c \tag{7.2a}$$

where c is the sound speed, $c = (\gamma p/\rho)^{1/2}$.

The corresponding right eigenvectors are

$$R_1(w) = \begin{pmatrix} 1 \\ u - c \\ H - uc \end{pmatrix} \quad , \quad R_2(w) = \begin{pmatrix} 1 \\ u \\ \frac{1}{2} u^2 \end{pmatrix} \quad , \quad R_3(w) = \begin{pmatrix} 1 \\ u + c \\ H + uc \end{pmatrix} \tag{7.2b}$$

where $H = (E + p)/\rho = c^2/(\gamma-1) + \frac{1}{2} u^2$ is the enthalpy.

Let $\alpha^k(w_L, w_R)$, $k = 1,2,3$, be the solution of the following system of linear equations (4.7)

$$w_R - w_L = \sum_{k=1}^{3} \alpha^k R^k(V(w_L,w_R)). \tag{7.3a}$$

where $V(w_L, w_R)$ (4.6) is some average state; denote its velocity and sound speed by $\hat{u}$ and $\hat{c}$, respectively. To calculate α^k in (7.3a) we first evaluate

$$C_1 = (\gamma-1) \{[E] + \frac{1}{2} \hat{u}^2[\rho] - \hat{u} [m]\}/\hat{c}^2 \tag{7.3b}$$

$$C_2 = \{[m] - \hat{u}[\rho]\}/\hat{c} \tag{7.3c}$$

where $[b]$ denotes $[b] = b_R - b_L$; then α^k in (7.3a) are obtained by

$$\alpha^1 = \frac{1}{2}(C_1 - C_2), \quad \alpha^2 = [\rho] - C_1, \quad \alpha^3 = \frac{1}{2}(C_1 + C_2). \tag{7.3d}$$

The second characteristic field corresponding to the eigenvalue u is linearly degenerate, i.e., $a_w^2 R^2 \equiv 0$; The other characteristic fields corresponding to the eigenvalues $u \pm c$ are genuinely nonlinear. Therefore in (6.1a) we define $\theta^2_{j+1/2}$ by (6.1b) for $k = 2$ and set $\theta^1_{j+1/2} = \theta^3_{j+1/2} \equiv 0$.

Given w_j^n we now list the operations needed to calculate w_j^{n+1}:

(i)   Compute: $(u_j = m_j/\rho_j)$ and (P_j) by (7.1c)

(ii)   Compute $\hat{u}_{j+1/2}$ and $\hat{c}_{j+1/2}$ from $V(v_j, v_{j+1})$; calculate $\max_j(|\hat{u}_{j+1/2}| + \hat{c}_{j+1/2})$; evaluate $\alpha^k_{j+1/2}$, $k = 1,2,3$ by (7.3b) – 7.3d).

Define $\lambda = \Delta t/\Delta x = \mu/\max_j(|\hat{u}_{j+1/2}| + \hat{c}_{j+1/2})$ where μ is the prescribed CFL restriction in (4.13)

(iii) Compute: $v^k_{j+1/2} = \lambda \hat{a}^k_{j+1/2}$ by (7.2a); $\tilde{g}^k_{j+1/2}$ by (4.8d);

$\theta^2_{j+1/2}$ by (6.1b) for $k = 2$; Set $\theta^1_{j+1/2} = \theta^3_{j+1/2} \equiv 0$.

(iv)   Calculate g^k_j by (6.1a).

(v)   Compute $\gamma^k_{j+1/2}$ by (4.8e), and $\bar{f}_{j+1/2}$ by (4.8b) and relations (7.2b)

(vi)   Compute w^{n+1}_j by (4.8a).

We note that in computing $\bar{f}_{j+1/2}$ in (v), one could take advantage of the simple form of the R^k in (7.2b).

Next we show how to implement Roe's linearization technique (4.14) – (4.15) in the above algorithm. Roe presents a particular form of averaging $V(w_L, w_R)$ such that for the Euler equations of gasdynamics, the mean value Jacobian $A(w_L, w_R)$ in (4.14a) can be expressed by (4.17). This averaging takes the following form:

$$\hat{u}_{j+1/2} = \langle \rho^{1/2} u \rangle / \langle \rho^{1/2} \rangle \quad , \quad \hat{H}_{j+1/2} = \langle \rho^{1/2} H \rangle / \langle \rho^{1/2} \rangle \quad ,$$

$$\tag{7.4a}$$

$$\hat{c}_{j+1/2} = \{(\gamma-1)(\hat{H}_{j+1/2} - \tfrac{1}{2}\hat{u}^2_{j+1/2})\}^{1/2}$$

where $\langle b \rangle$ denotes the arithemetic mean

$$\langle b \rangle = \tfrac{1}{2}(b_j + b_{j+1}) \tag{7.4b}$$

Therefore to use Roe's linearization in our scheme all one has to do is to compute $\hat{u}_{j+1/2}$ and $\hat{c}_{j+1/2}$ in (ii) by (7.4).

We remark that the averaging in (7.4) is rather expensive. It seems to us that in many applications the simple arithmethic average (7.4b) will do just as well.

8. Numerical experiments.

In this section we present some numerical experiments that demonstrate the performance of the proposed second order accurate scheme. We consider here the following two versions of it:

$$v_j^{n+1} = v_j^n - \lambda(\bar{f}_{j+1/2} - \bar{f}_{j-1/2}) \tag{8.1a}$$

$$\bar{f}_{j+1/2} = \frac{1}{2}[f(v_j) + f(v_{j+1}) - \frac{1}{\lambda}\sum_{k=1}^{m} \beta_{j+1/2}^k R_{j+1/2}^k] \tag{8.1b}$$

where

$$\beta_{j+1/2}^k = Q^k(\nu_{j+1/2}^k + \gamma_{j+1/2}^k)a_{j+1/2}^k - (g_j^k + g_{j+1}^k) \tag{8.2}$$

and g_j^k is defined by an appropriate variant of (6.1) and (4.8d); $\gamma_{j+1/2}^k$ is defined by (4.8e).

The first version is (8.1) - (8.2) with

$$Q^k(x) \equiv |x| \tag{8.3a}$$

and will be referred to as the scheme ULT1; The second version is (8.1) - (8.2) with

$$Q^k(x) \equiv x^2 + 1/4 \tag{8.3b}$$

and will be referred to as the scheme ULT2.

For comparison sake we also present calculations with the following two schemes:

(i)   The second order accurate Lax-Wendroff-type scheme (8.1) with

$$\beta_{j+1/2}^k = (\nu_{j+1/2}^k)^2 \tag{8.4}$$

which is referred to as the LW scheme.

(ii) The first order accurate Godunov-type scheme of Roe (see [15] and [10])
which is defined by (8.1) with

$$\beta_{j+1/2}^k = |\nu_{j+1/2}^k| \tag{8.5}$$

and is referred to as the ROE scheme.

       In all the schemes and experiments reported herein we use the Roe
linearization (7.3) - (7.4).

(I)   The shock tube problem.
       We cosider now a Riemann problem

$$w(x,0) = \begin{cases} w_L & x < 0 \\ w_R & x > 0 \end{cases} \tag{8.6a}$$

for the Euler equations of a polytropic gas (7.1). Our first set of data is

$$W_L = \begin{pmatrix} 0.445 \\ 0.3111 \\ 8.928 \end{pmatrix} \qquad , \qquad W_R = \begin{pmatrix} 0.5 \\ 0 \\ 1.4275 \end{pmatrix} \tag{8.6b}$$

Other numerical experiments with this problem are reported in [6] and the references cited there.

In figures 2a, 2b and 2c we show the results obtained by the ROE, LW and ULT1 schemes, respectively. The numerical values are shown by circles; the exact solution is shown by the solid line. The calculations in figure 2 were performed with 100 time-steps under the CFL restriction $\mu = 0.95$ in (4.13), and 140 cells.

In figure 3 we repeat the calculations presented in figure 2 for a different set of data for the Riemann Problem (8.6a)

$$W_L = \begin{Bmatrix} 1. \\ 0. \\ 2.5 \end{Bmatrix} \quad , \quad W_R = \begin{Bmatrix} 0.125 \\ 0. \\ 0.25 \end{Bmatrix}$$

Other numerical experiments with this problem are presented in [16]. The calculations in figure 3 were performed with 50 time-steps under the CFL restriction $\mu = 0.95$ in (4.13), with 100 cells.

We remark that the solution of ULT2 for these probelms look almost identical to those of ULT1. In both schemes we find almost no dependence on the CFL number.

(II) The Quasi 1-D nozzle problem.

We consider an axisymmetric nozzle with a cross-section area $A(x)$. The cross-section average of the flow satisfy the following one-dimensional system of equations

$$w_t + f(w)_x = - s(w,x) \quad , \quad s(w,x) = \begin{Bmatrix} 0 \\ p \dfrac{dA}{dx} \\ 0 \end{Bmatrix}$$

where w, $f(w)$ and p are given in (7.1).

In figures 4,5 and 6 we[2] present numerical approximations to steady state solutions of (8.7a).

In figures 4 and 5 we show solutions for a divergent nozzle with the cross-section area

$$A(x) = 1.398 + 0.347 \quad \tanh (0.8x - 4); \tag{8.7b}$$

the flow condition is supersonic at the entrance and subsonic at the exit. Figures 4a and 4b show steady state solutions on a crude mesh of the ROE and ULT1 schemes, respectively. Figure 5 shows the ULT1 results for the same problem on a finer mesh.

In figure 6 we show a steady state solution of the ULT1 scheme for a convergent-divergent nozzle with the cross-section area

$$A(x) = \begin{cases} 1 + (A_0 - 1)(1 - x/5)^2 & x \leqslant 5 \\ 1 + (A_E - 1)[(x - 5)/(x_E - 5)]^2 & x > 5 \end{cases} \tag{8.7c}$$

where A_0 = entrance area, A_E = exit area; Here the flow is subsonic at the entrance as well as at the exit.

---

2   These figures are by courtesy of Helen C. Yee of the NASA-Ames Research Center.

The exact solutions in figures 4 to 6 are shown by the solid curve; the values of the numerical solutions are indicated by a rombus.

Hopefully a detailed report of these calculations and the particular approach to steady state will be published elsewhere.

(III) 2-D Flow through a duct.

In figures 7a and 7b we show solutions to the problem of the flow of air through a duct containing a step. Initially the flow is everywhere to the right at Mach 3, with $\rho = 1.4$, $p = 1$ and $c = 1$. The duct width is 1, its length is 3, and the step of hieght 0.2 is located a distance of 0.6 from the entrance. Figure 7 shows the results at $t = 4$ with a crude uniform Cartesian grid with $\Delta x = \Delta y = 0.1$.

This problem was used by Woodward and Colella to test the performance of various numerical schemes (see [18] and the references cited htere).

The solutions in figure 7 were obtained by a Strang -type dimensional splitting of the form

$$v^{n+2} = L v^n \tag{8.8a}$$

$$L = L_x \, L_y \, L_y \, L_x \tag{8.8b}$$

where L_x and L_y are one-dimenional finite difference operators approximating

$$L_x : w_t + f(w)_x = 0 \quad , \quad L_y: w_t + g(w)_y = 0. \tag{8.8c}$$

If L_x and L_y are stable and dissipative second order accurate approximations to the one-dimensional equations in (8.8c), then the scheme (8.8a) - (8.8b) is a stable second order accurate approximation to the 2-dimensional problem.

$$w_t + f(w)_x + g(w)_y = 0. \tag{8.8d}$$

In figure 7a we show for comparison sake the results of the second order accurate hybrid scheme (see [6]) (8.1) with

$$\beta^k_{j+1/2} = [(\upsilon^k_{j+1/2})^2 + \frac{1}{4} \theta^k_{j+1/2}] \, \alpha^k_{j+1/2} \tag{8.9}$$

where $\theta^k_{j+1/2}$ is (6.1b) for all k.

In figure 7b we show the results of ULT2.

Both figures 7a and 7b show 30 equally spaced density contours; Both calculations were performed with a CFL restriction of 0.75

We remark that the corner of the step in our calculations is treated as a sharp corner without any rounding (or equivalent addition of numerical viscosity). The sonic line emerging from this corner is a curve on which an eigenvalue of (8.8d) vanishes. It is interesting to note that the results of the Godunov scheme in figure 1a of [18] indicate that part of the sonic line may turn into an expansion shock. We find a similar behaviour in the results of ULT1 for this problem.

Altogether we find the performance of the new second order accurate scheme to be quite pleasing. We note that the scheme is simple to program and requires only slightly more CPU time than a Lax-Wendroff scheme with some artificial viscosity.

Acknowledgement.

References.

[1] J.P. Boris and D.L.Book, J. Computational Phys., 11 (1973), pp. 38-69.

[2] M.G.Crandall and A Majda, Math. Comp., 34 (1980), pp. 1-21.

[3] J. Glimm, Comm. Pure Appl. Math., 18 (1965), pp. 697 - 715.

[4] S.K. Godunov, Math. Sbornik, 47 (1959), pp. 271-295; also: Cornell
     Aero. Lab. Transl.

[5] A. Harten, Comm. Pure Appl. Math., 30 (1977). pp. 611-638.

[6] A. Harten, Math. Comp., 32 (1978), pp. 363-389.

[7] A. Harten, "On the symmetric form of systems of conservation laws with
     entropy", ICASE Report 1981, to appear.

[8] A. Harten, J.M. Hyman and P.D.Lax.,Comm. Pure Appl. Math., 29 (1976),
     pp. 297-322.

[9] A.Harten and P.D.Lax, SIAM J. Num. Anal., 18 (1981), pp.289-315.

[10] A. Harten, P.D.Lax and B. van Leer, "On upstream differencing and Godunov-
     type schemes for hyperbolic conservation laws", ICASE Report, to appear.

[11] P.D. Lax, "Hyperbolic systems of conservation laws and the mathematical
     theory of shock waves", SIAM, Philadelphia, 1972.

[12] T.P. Liu, J. Math. Anal. Appl., 53 (1976), pp. 73-88.

[13] S. Osher, "Numerical solution of singular perturbation problems and hyper-
     bolic systems of conservation laws", UCLA Reprint 1980, to appear.

[14] P.L. Roe, Proc. Seventh Int. Conf. Num. Meth. Fluid Syn., Stanford/ NASA
     Ames, June 1980, Springer-Verlag.

[15] P.L. Roe, "Approximate Riemann solvers, parameter vectores and difference
     schemes", J Computational Phys., to appear.

[16] G.A. Sod, J. Computational Phys., 27 (1978), pp. 1-31.

[17] B. van Leer, J. Computational Phys., 14 (1974), pp. 361-370.

[18] P. Woodward and P. Colella, Proc. Seventh Int. Conf. Num. Meth. Fluid Dyn.,
     Stanford/NASA Ames, June 1980, Springer-Verlag.
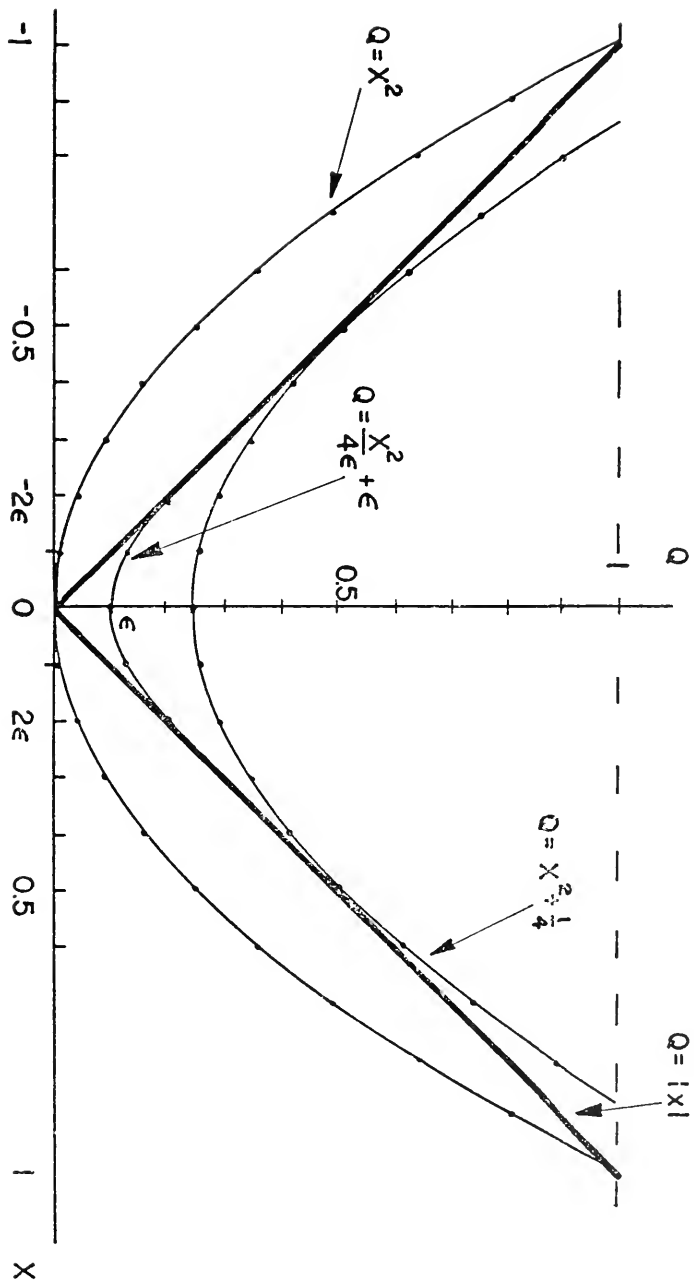
Figure 1.
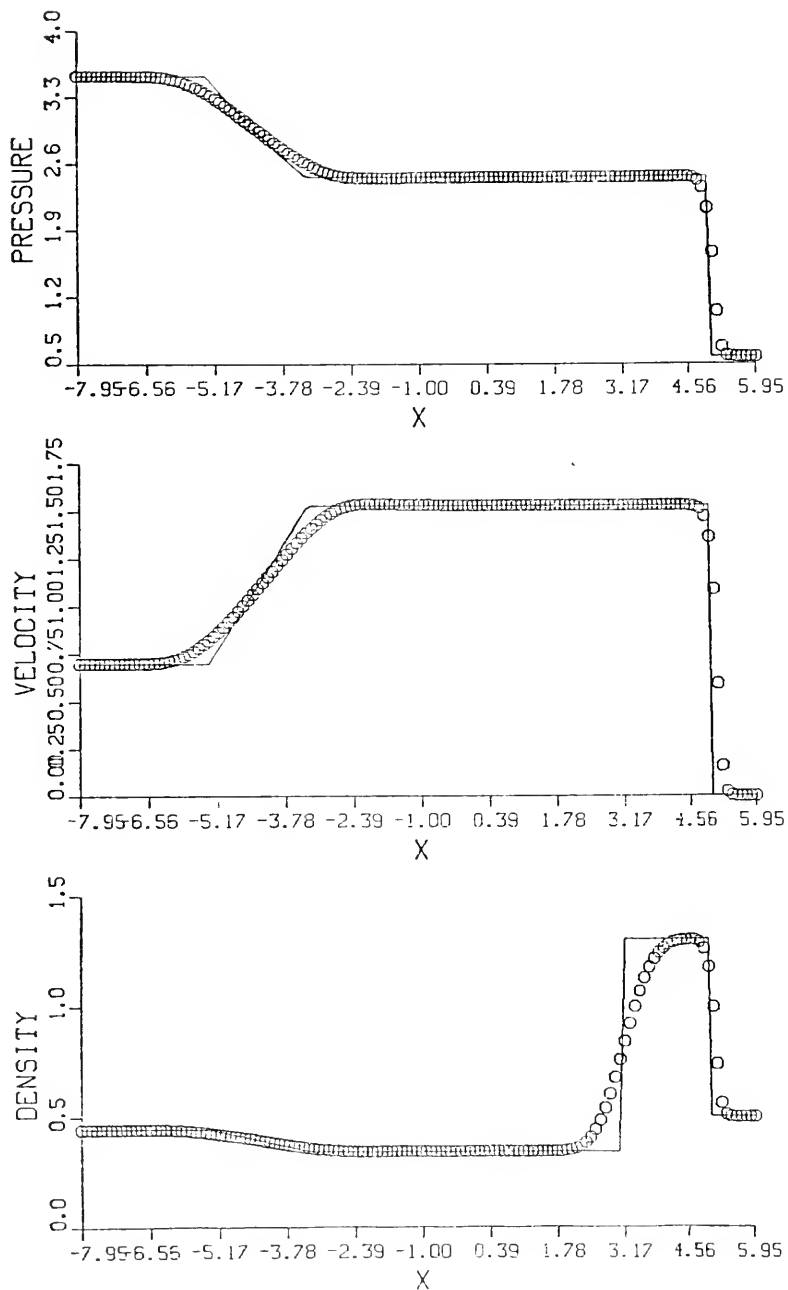
Figure 2a. ROE scheme for (8.6b)
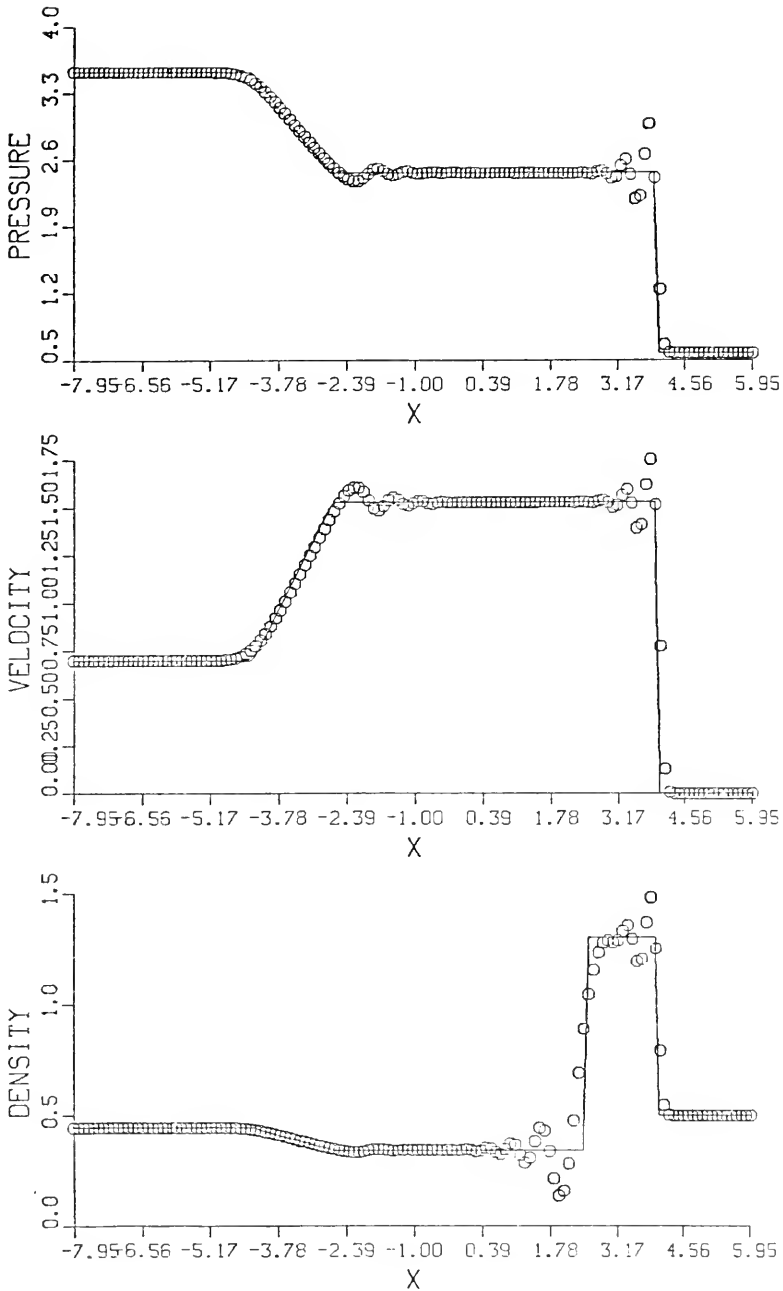
Figure 2b. LW scheme for (8.6b)
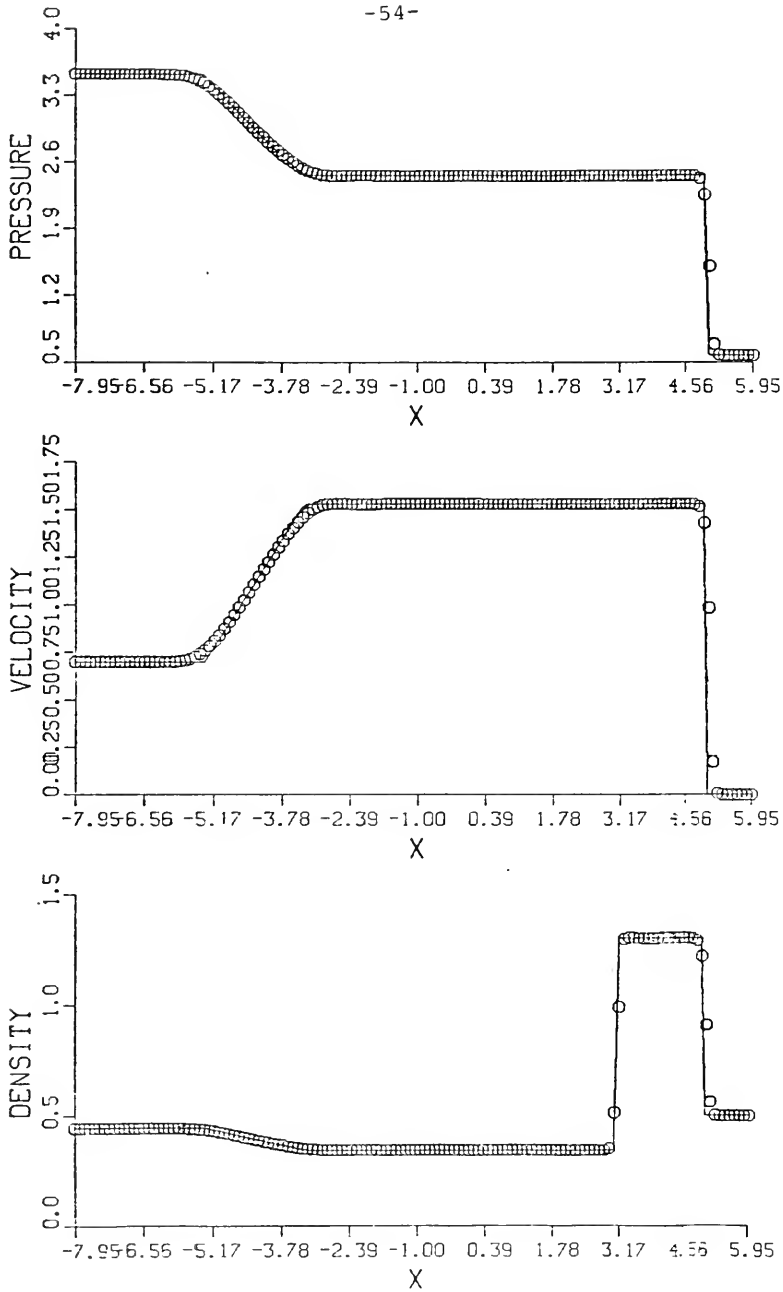
Figure 2c. ULT1 scheme for (8.6b)
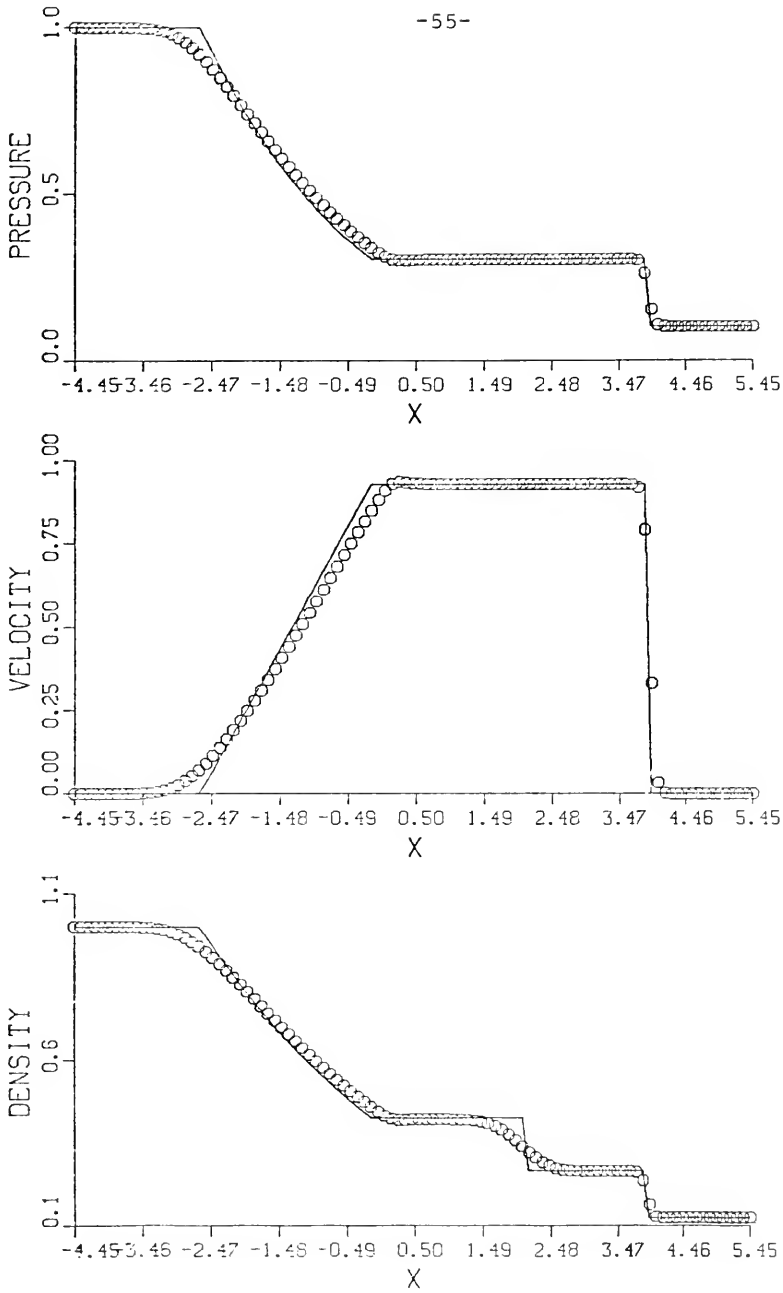
Figure 3a. ROE scheme for (8.6c)
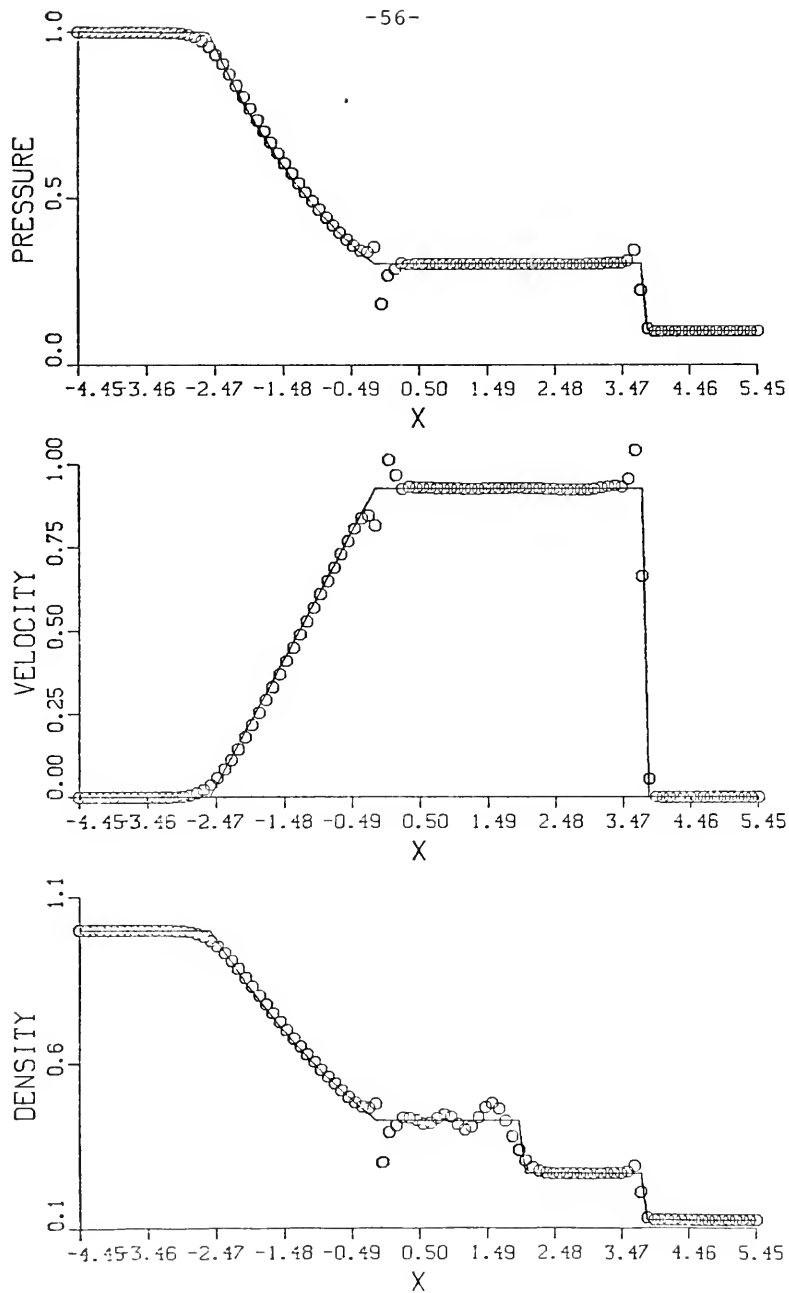
Figure 3b. LW scheme for (8.6c)
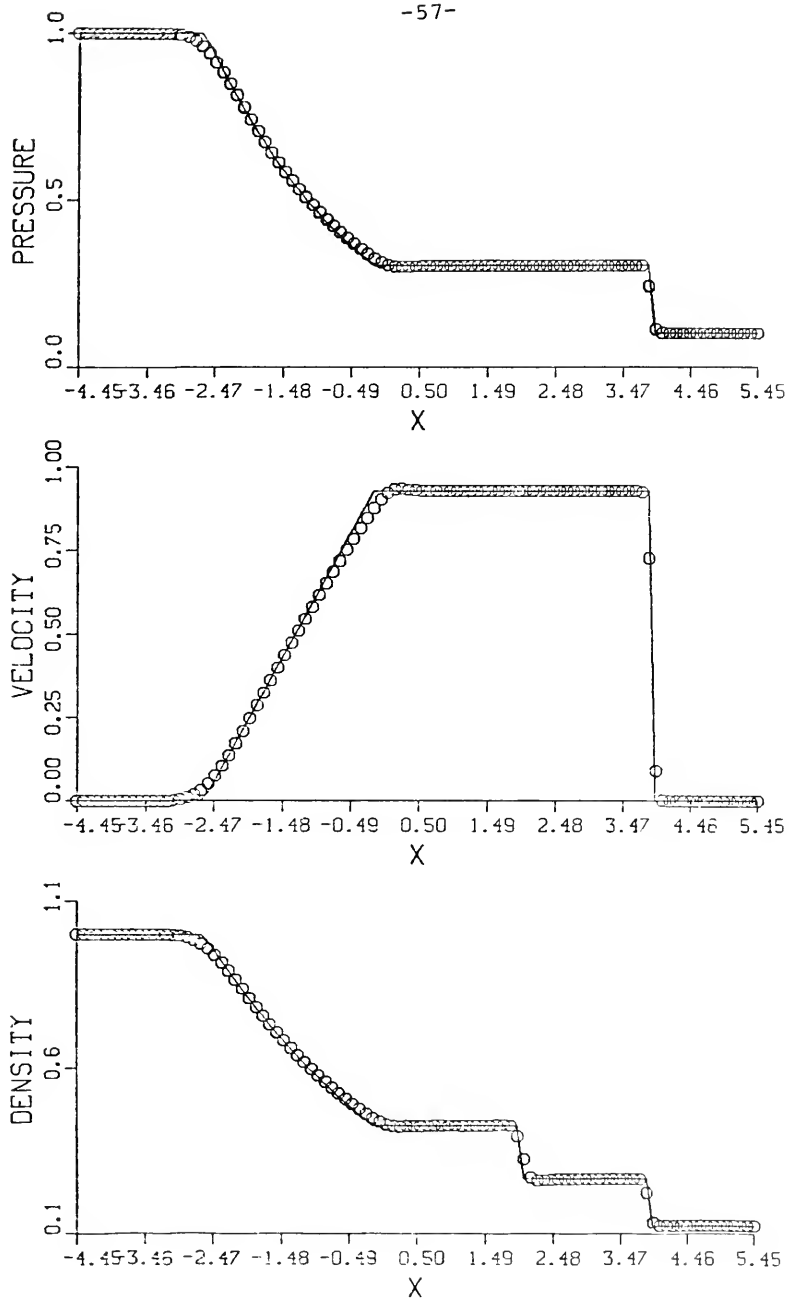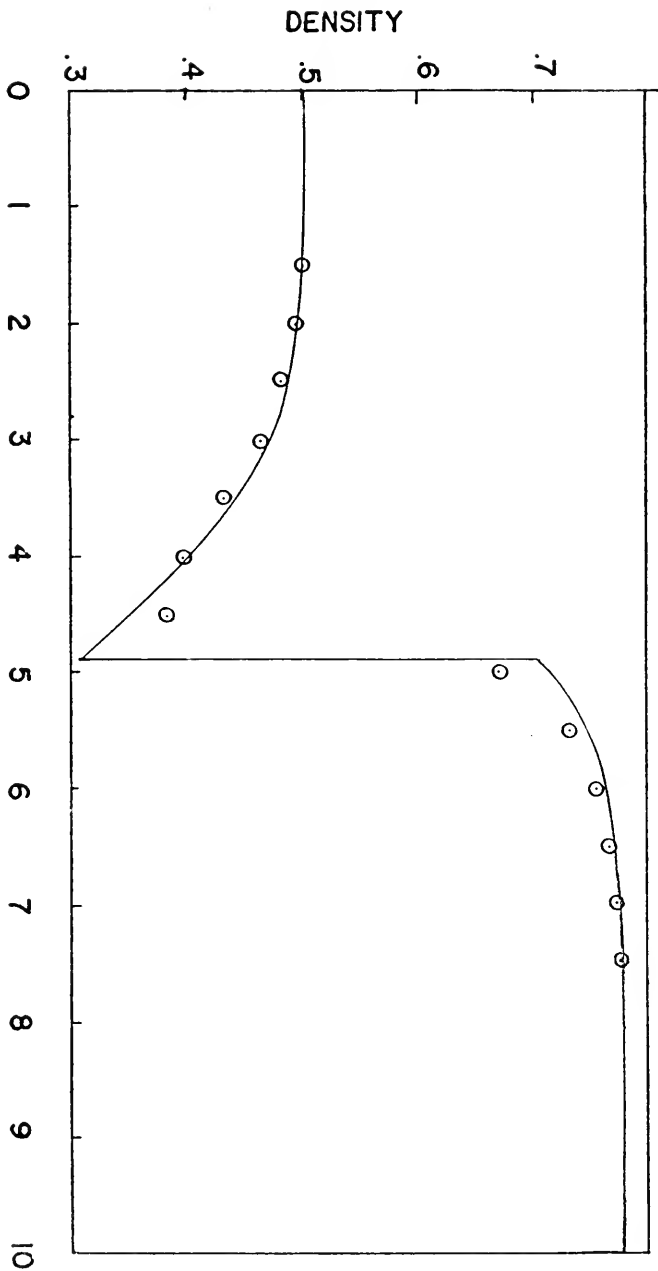
Figure 3c. ULT1 scheme for (8.6c)
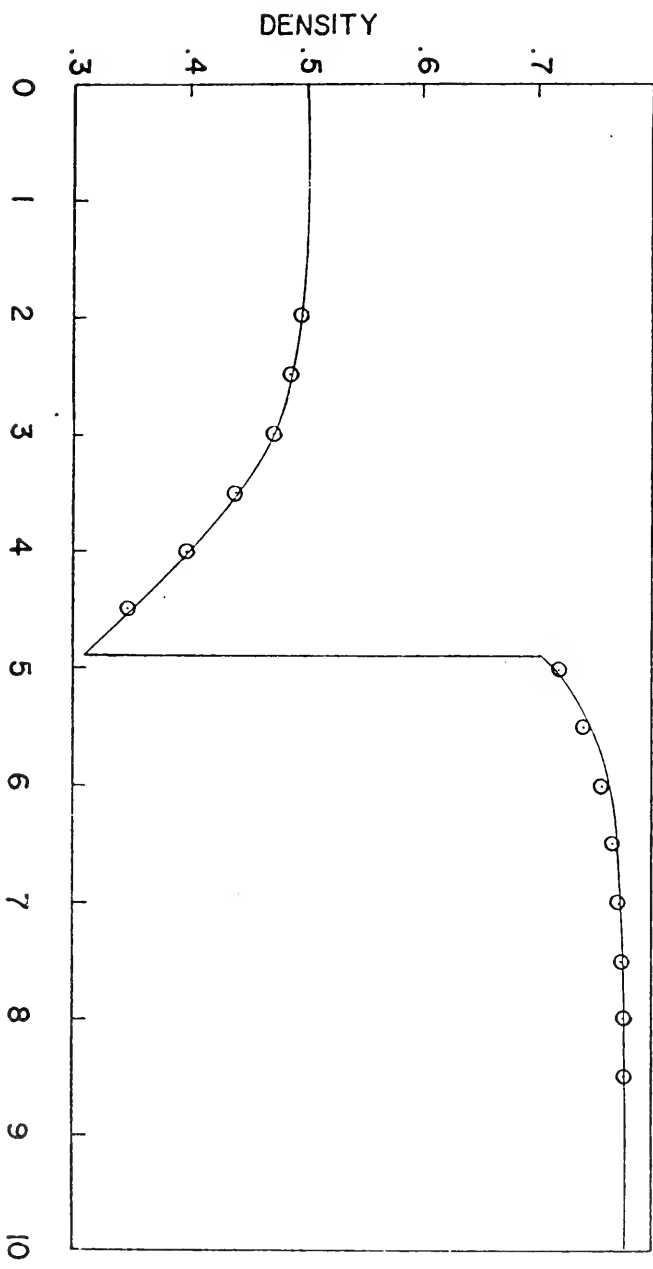
Figure 4a. ROE scheme for (8.7b).

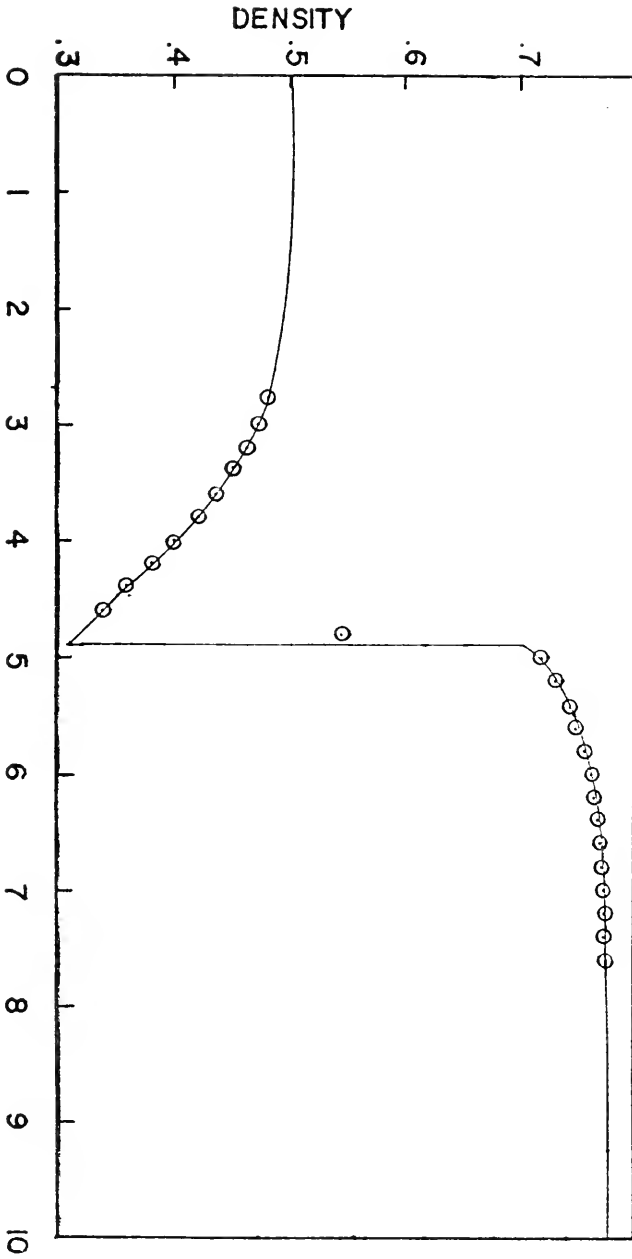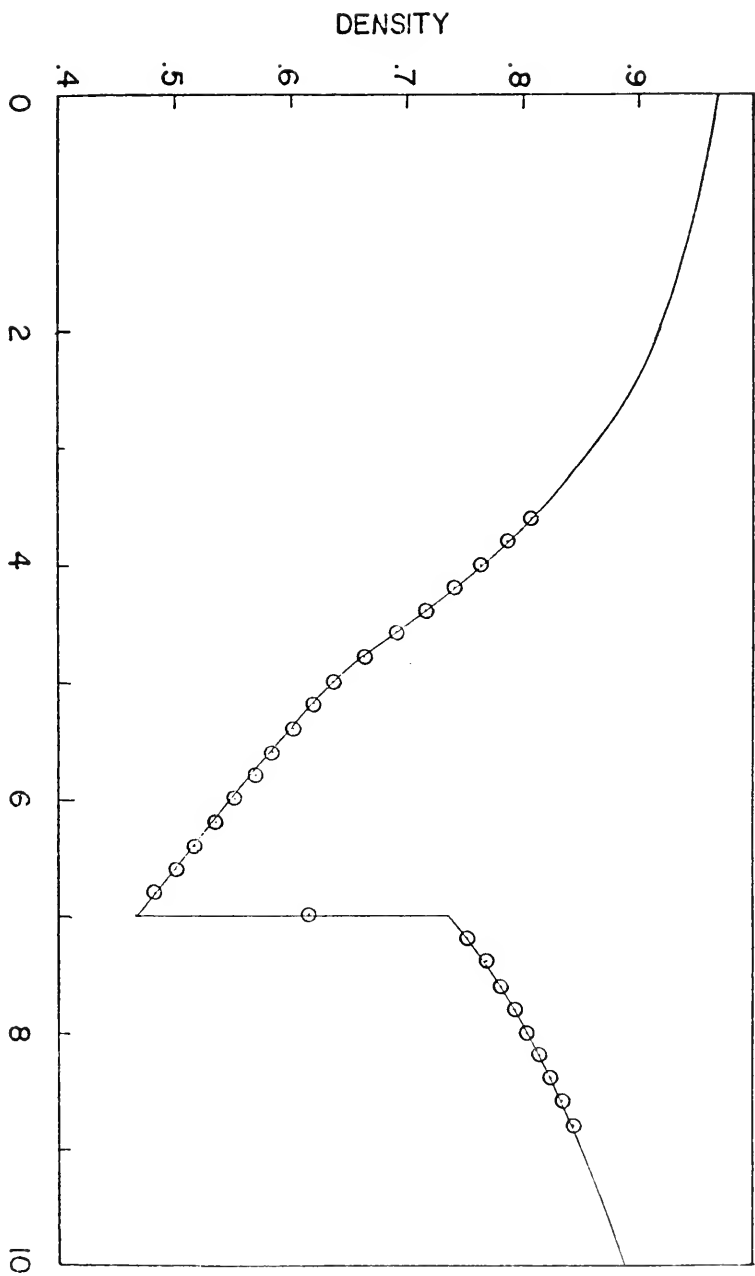Figure 4b. ULTI scheme for (8.7b).

Figure 5. ULTI scheme for (8.7b) - finer grid.

Figure 6. ULT1 scheme for (8.7c).
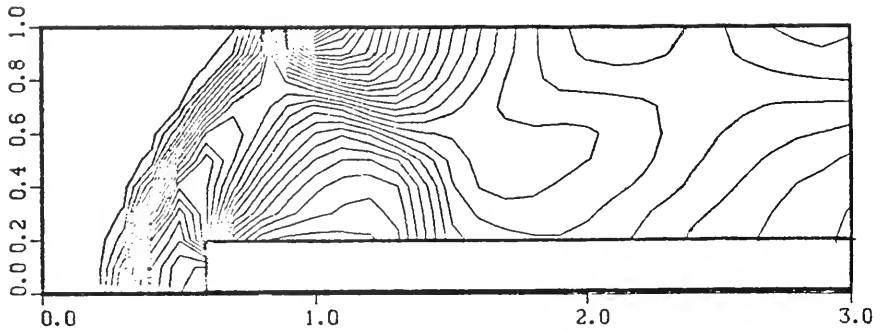
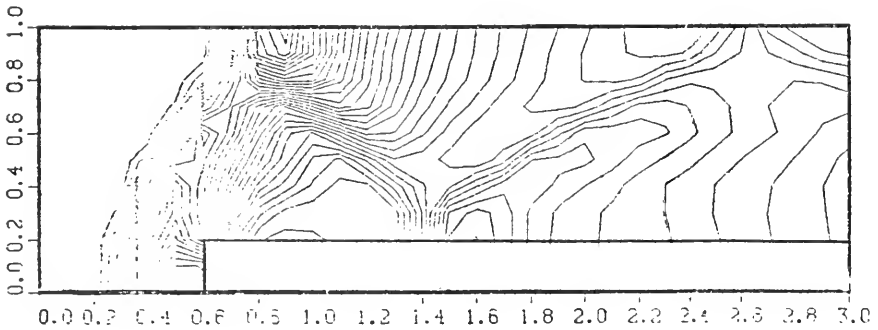Figure 7a. Hybrid scheme [(8.1) + (8.9)] – 10 x 30 grid.

Fig. 7b. ULT2 scheme – 10 x 30 grid.

N.Y.U. Courant Institute of
Mathematical Sciences
251 Mercer St.
New York, N. Y. 10012

This book may be kept   1362

FOURTEEN DAYS

A fine will be charged for each day the book is kept overtime.